

Fundamentos de Estadística para las Ciencias de la Vida

4.ª edición

Myra L. Samuels
Jeffrey A. Witmer
Andrew Schaffner

booksmedicos.org

Fundamentos de Estadística para las Ciencias de la Vida

booksmedicos.org

Fundamentos de Estadística para las Ciencias de la Vida

4.^a edición

Myra L. Samuels
Purdue University

Jeffrey A. Witmer
Oberlin College

Andrew A. Schaffner
*California Polytechnic State University,
San Luis Obispo*

Traducción
Javier Portillo García
Profesor Titular de Universidad
Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad Politécnica de Madrid

Revisión técnica
Martín Cera López
Departamento de Matemática Aplicada I
Universidad de Sevilla

booksmedicos.org

PEARSON

Datos de catalogación bibliográfica

Fundamentos de Estadística para las Ciencias de la Vida, 4.^a edición
Myra L. Samuels, Jeffrey A. Witmer, Andrew A. Schaffner

PEARSON EDUCACIÓN, S.A., Madrid, 2012

ISBN: 9788478291373

Materia: 311. Teoría de la estadística

Formato: 195 × 250 mm Páginas: 648

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser utilizada con la autorización de sus titulares, salvo excepción prevista por la ley. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (arts. 270 y sgts. Código penal).

Diríjase a CEDRO (Centro Español de Derechos Reprográficos –www.cedro.org), si necesita fotocopiar o escanear algún fragmento de esta obra.

Todos los derechos reservados.

© 2012, PEARSON EDUCACIÓN, S.A.

C/ Ribera del Loira, 28

28042 Madrid (España)

www.pearson.es

Authorized translation from the English language edition, entitled STATISTICS FOR THE LIFE SCIENCES, 4th Edition by MYRA SAMUELS; JEFFREY WITMER; ANDREW SCHAFFNER.

All rights reserved. No part of this book may be reproduced or transmitted in any form or any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

SPANISH language edition published by Pearson Educación, S.A., Copyright © 2012.

ISBN: 9788478291373

Depósito Legal: M-10290-2012

Equipo de edición:

Editor: Miguel Martín-Romo

Técnico editorial: Esther Martín

Equipo de diseño:

Diseñadora: Elena Jaramillo

Técnico de diseño: Irene Medina

Equipo de producción:

Directora de producción: Marta Illescas

Coordinadora de producción: Tini Cardoso

Diseño de cubierta: Copibook, S.L.

Composición: Copibook, S.L.

Impreso por:

IMPRESO EN ESPAÑA - PRINTED IN SPAIN

Nota sobre enlaces a páginas web ajenas: este libro incluye enlaces a sitios web cuya gestión, mantenimiento y control son responsabilidad única y exclusiva de terceros ajenos a PEARSON EDUCACIÓN, S.A. Los enlaces u otras referencias a sitios web se incluyen con finalidad estrictamente informativa y se proporcionan en el estado en que se encuentran en el momento de publicación sin garantías, expresas o implícitas, sobre la información que se proporcione en ellas. Los enlaces no implican el aval de PEARSON EDUCACIÓN S.A. a tales sitios, páginas web, funcionalidades y sus respectivos contenidos o cualquier asociación con sus administradores. En consecuencia, PEARSON EDUCACIÓN S.A., no asume responsabilidad alguna por los daños que se puedan derivar de hipotéticas infracciones de los derechos de propiedad intelectual y/o industrial que puedan contener dichos sitios web ni por las pérdidas, delitos o los daños y perjuicios derivados, directa o indirectamente, del uso de tales sitios web y de su información. Al acceder a tales enlaces externos de los sitios web, el usuario estará bajo la protección de datos y políticas de privacidad o prácticas y otros contenidos de tales sitios web y no de PEARSON EDUCACIÓN S.A.

Este libro ha sido impreso con papel y tintas ecológicos

CONTENIDO

Prólogo	ix
1 Introducción	1
1.1 La Estadística y las Ciencias de la Vida	1
1.2 Tipos de evidencia	7
1.3 Muestreo aleatorio	14
Notas	24
2 Descripción de muestras y poblaciones	26
2.1 Introducción	26
2.2 Distribuciones de frecuencias	28
2.3 Estadística descriptiva: medidas de tendencia central	39
2.4 Diagramas de caja	44
2.5 Relaciones entre variables	51
2.6 Medidas de dispersión	58
2.7 Efectos de transformación de variables (opcional)	66
2.8 Inferencia estadística	71
2.9 Perspectiva	76
Notas	81
3 La probabilidad y la distribución binomial	84
3.1 La probabilidad y las Ciencias de la Vida	84
3.2 Introducción a la probabilidad	84
3.3 Las reglas de la probabilidad (opcional)	94
3.4 Curvas de densidad	98
3.5 Variables aleatorias	102
3.6 La distribución binomial	106
3.7 Ajuste de una distribución binomial a los datos (opcional)	114
Notas	120
4 La distribución normal	122
4.1 Introducción	122
4.2 Las curvas normales	124

4.3 Áreas bajo una curva normal	126
4.4 Evaluación de la normalidad	133
4.5 Perspectiva	143
Notas	146
5 Distribuciones muestrales	147
5.1 Ideas básicas	147
5.2 La media muestral	151
5.3 Ilustración del Teorema Central del Límite (opcional)	161
5.4 La aproximación normal de la distribución binomial (opcional)	164
5.5 Perspectiva	170
Notas	171
6 Intervalos de confianza	173
6.1 Estimación estadística	173
6.2 Error típico de la media	174
6.3 Intervalo de confianza para μ	179
6.4 Planificación de un estudio para estimar μ	190
6.5 Condiciones de validez de los métodos de estimación	192
6.6 Comparación de dos medias	201
6.7 Intervalo de confianza para $(\mu_1 - \mu_2)$	207
6.8 Perspectiva y resumen	213
Notas	218
7 Comparación de dos muestras independientes	221
7.1 Contraste de hipótesis: el test de aleatorización	221
7.2 Contraste de hipótesis: el test t	226
7.3 Comentarios adicionales sobre el test t	237
7.4 Asociación y causalidad	244
7.5 Tests t de una cola	251
7.6 Más sobre la interpretación de la significación estadística	260
7.7 Planificación de la potencia adecuada (opcional)	267
7.8 t de Student: condiciones y resumen	272
7.9 Más sobre los principios del contraste de hipótesis	276
7.10 El test de Wilcoxon-Mann-Whitney	280
7.11 Perspectiva	289
Notas	296

8 Comparación de dos muestras pareadas	300
8.1 Introducción	300
8.2 El intervalo de confianza y el test t para muestras pareadas	301
8.3 El diseño pareado	311
8.4 El test de signos	316
8.5 El test de rangos con signo de Wilcoxon	322
8.6 Perspectiva	327
Notas	335
9 Datos en categorías: distribuciones para una muestra	337
9.1 Observaciones dicotómicas	337
9.2 Intervalo de confianza de la proporción poblacional	342
9.3 Otros niveles de confianza (opcional)	348
9.4 Inferencia sobre proporciones: el test chi-cuadrado de bondad de ajuste	349
9.5 Perspectiva y resumen	359
Notas	363
10 Datos en categorías: relaciones	366
10.1 Introducción	366
10.2 El test chi-cuadrado para la tabla de contingencia de 2×2	368
10.3 Independencia y asociación en la tabla de contingencia de 2×2	376
10.4 Test exacto de Fischer (opcional)	384
10.5 La tabla de contingencia de $r \times k$	389
10.6 Aplicabilidad de los métodos	394
10.7 Intervalo de confianza de la diferencia de probabilidades	397
10.8 Datos pareados y tablas de 2×2 (opcional)	400
10.9 Riesgo relativo y razón de ventajas (opcional)	402
10.10 Resumen del test chi-cuadrado	410
Notas	415
11 Comparación de medias de varias muestras independientes	419
11.1 Introducción	419
11.2 El análisis de la varianza con un factor	423
11.3 Análisis del modelo ANOVA	431
11.4 El test F global	433
11.5 Aplicabilidad de los métodos	437
11.6 Diseño de bloques aleatorizados con un factor	441

11.7 ANOVA de dos factores	452
11.8 Combinaciones lineales de medias (opcional)	459
11.9 Comparaciones múltiples (opcional)	468
11.10 Perspectiva	477
Notas	482
12 Regresión lineal y correlación	485
12.1 Introducción	485
12.2 El coeficiente de correlación	487
12.3 La recta de regresión	497
12.4 Interpretación paramétrica de la regresión: el modelo lineal	510
12.5 Inferencia estadística con respecto a β_1	516
12.6 Pautas para la interpretación de la regresión y de la correlación	521
12.7 Precisión de la predicción (opcional)	532
12.8 Perspectiva	535
12.9 Resumen de fórmulas	546
Notas	554
13 Resumen de métodos de inferencia	557
13.1 Introducción	557
13.2 Ejemplos de análisis de datos	559
Notas	572
Apéndices	573
Tablas estadísticas	587
Soluciones de ejercicios seleccionados	615
Índice alfabético	623
Índice de ejemplos	630

PRÓLOGO

Fundamentos de Estadística para las Ciencias de la Vida es un texto introductorio sobre estadística, dirigido especialmente a estudiantes que se están especializando en Ciencias de la Vida. Sus principales objetivos son: (1) mostrar a los estudiantes cómo se utiliza el razonamiento estadístico en investigación biológica, médica y agrícola; (2) permitir a los estudiantes realizar con confianza análisis estadísticos simples e interpretar los resultados, y (3) despertar la conciencia de los estudiantes sobre cuestiones estadísticas básicas como la aleatorización, la confusión y el papel de las réplicas independientes.

Estilo y enfoque

El estilo de *Fundamentos de Estadística para las Ciencias de la Vida* es informal y utiliza solo la notación matemática mínima. No existen prerequisites excepto álgebra elemental; cualquier persona que haya leído un libro de texto sobre biología o química puede leer este texto. Es adecuado para su uso por estudiantes graduados o pregraduados en biología, agronomía, ciencias médicas y de la salud, nutrición, farmacia, veterinaria, educación física, ciencias forestales y otras ciencias de la vida.

Uso de datos reales Los ejemplos reales son más interesantes y a menudo más aclaratorios que los artificiales. *Fundamentos de Estadística para las Ciencias de la Vida* incluye cientos de ejemplos y ejercicios que utilizan datos reales, que representan una amplia variedad de los temas de investigación en Ciencias de la Vida. Cada ejemplo se ha elegido para ilustrar una cuestión estadística concreta. Los ejercicios se han diseñado para reducir el esfuerzo de cálculo y centrar la atención de los estudiantes en los conceptos y las interpretaciones.

Énfasis en las ideas El texto enfatiza las ideas estadísticas en vez de los cálculos o las formulaciones matemáticas. La teoría de la probabilidad se presenta solo para dar soporte a los conceptos estadísticos. En las presentaciones de estadística descriptiva e inferencial, se hace hincapié en la interpretación. Mediante ejemplos destacados, se muestra al estudiante por qué es importante que un análisis sea apropiado para el problema de investigación que se considera, para el diseño estadístico del estudio y para la naturaleza de las distribuciones subyacentes. Se advierte al estudiante sobre el error de confundir la falta de significación estadística con la insignificancia práctica y se fomenta el uso de los intervalos de confianza para valorar la magnitud de un efecto. Se conduce al estudiante para que reconozca el impacto sobre la investigación real de los conceptos de diseño como el muestreo aleatorio, la aleatorización, la eficiencia y el control de variaciones externas mediante la agrupación en bloques o el ajuste. Numerosos ejercicios amplifican y refuerzan la captación de estas ideas por los estudiantes.

El papel de la tecnología El análisis de datos de investigación se realiza generalmente con la ayuda de un computador. En diversas partes del texto se muestran gráficos generados por computador. Sin embargo, en el estudio de la estadística es deseable que el estudiante gane experiencia trabajando directamente con los datos y utilizando lápiz, papel y una calculadora de mano, además del computador. Esta experiencia ayudará al estudiante a apreciar la naturaleza y el propósito de los cálculos estadísticos. El estudiante estará así preparado para realizar un uso inteligente del computador, es decir, proporcionarle instrucciones adecuadas e interpretar adecuadamente los resultados. De acuerdo con esto, la mayoría de los ejercicios de este texto se pueden resolver a mano. También se proporcionan los archivos electrónicos de datos para muchos ejercicios, de manera que, si se desea, se pueda utilizar el computador. Algunos ejercicios seleccionados se identifican como candidatos a ser resueltos con ayuda de un computador. Generalmente, este tipo de ejercicios requieren cálculos excesivamente largos para realizarlos a mano.

Organización

Este texto está organizado de forma que permite tratar en un semestre el máximo número de ideas estadísticas importantes, incluyendo la potencia, la inferencia múltiple y los principios básicos de diseño. Incluyendo o excluyendo las secciones opcionales, el profesor puede utilizar también el texto para un curso de uno o dos cuatrimestres. Es adecuado tanto para un curso final como para el primer curso de una secuencia.

A continuación se presenta un breve resumen del texto.

Capítulo 1: Introducción. La naturaleza y el impacto de la variabilidad en datos biológicos. Los peligros de los estudios observacionales, en contraste con experimentos. El muestreo aleatorio.

Capítulo 2: Descripción de distribuciones. Distribuciones de frecuencias, estadística descriptiva, concepto de población frente a muestra.

Capítulos 3, 4 y 5: Preparación teórica. Probabilidad, distribuciones binomial y normal, distribuciones muestrales.

Capítulo 6: Intervalos de confianza de la media y de la diferencia de medias.

Capítulo 7: Contraste de hipótesis, con énfasis en el test t . El test de aleatoriedad, el test de Wilcoxon-Mann-Whitney.

Capítulo 8: Inferencia para muestras pareadas. Intervalo de confianza. Test t , test de signos y test de rangos con signo de Wilcoxon.

Capítulo 9: Inferencia sobre la proporción. Intervalos de confianza y test chi-cuadrado de bondad de ajuste.

Capítulo 10: Relaciones en datos en categorías. Probabilidad condicional, tablas de contingencia. Las secciones opcionales cubren el test exacto de Fischer, el test de McNemar y las razones de ventaja.

Capítulo 11: Análisis de la varianza. Diseño con un factor, procedimientos de comparación múltiple, ANOVA en bloques con un factor, ANOVA con dos factores. Las secciones opcionales incluyen contrastes y comparaciones múltiples.

Capítulo 12: Correlación y regresión. Aspectos descriptivos y deductivos sobre la regresión lineal y la correlación entre ambas.

Capítulo 13: Resumen de métodos de inferencia.

Al final del libro se proporcionan tablas estadísticas. Las tablas de valores críticos son especialmente sencillas de utilizar, porque siguen disposiciones mutuamente consistentes y, por tanto, se utilizan de la misma manera.

Los apéndices opcionales del final del libro proporcionan al estudiante interesado una visión más profunda de temas como por ejemplo el cálculo de la distribución nula en el test de Wilcoxon-Mann-Whitney.

Cambios en la cuarta edición

- Parte del material que estaba anteriormente en el Capítulo 8, sobre principios estadísticos de diseño, se encuentra ahora en el Capítulo 1. Otras partes del anterior Capítulo 8 se encuentran ahora repartidas por el libro, con la intención de que los estudiantes lleguen a apreciar que en todos los estudios estadísticos intervienen aspectos como la recogida de datos y el ámbito de la inferencia (una buena parte de las gráficas no se estudian y utilizan de forma aislada sino que son una parte central del análisis estadístico y aparecen por tanto por todo el libro).
- Se han reorganizado algunos otros capítulos. Entre los cambios se incluyen:
 - La inferencia sobre una sola proporción se ha trasladado desde el Capítulo 6 al nuevo Capítulo 9.
 - El intervalo de confianza para la diferencia de medias se ha trasladado desde el Capítulo 7 al Capítulo 6.

- Un nuevo capítulo (9) presenta los procedimientos de inferencia para una variable categórica de la que se observa una sola muestra.
- El Capítulo 11 incluye un tratamiento más profundo del ANOVA de dos factores y de procedimientos de comparación múltiple en el análisis de la varianza.
- El Capítulo 12 empieza ahora con la correlación y sigue después con la regresión, en lugar de hacerlo al revés.
- El 25 % de los problemas del libro son nuevos o han sido revisados. Como anteriormente, la mayoría se basan en datos reales y están escogidos de diversas materias de interés dentro de las principales asignaturas de las Ciencias de la Vida. Los conjuntos de datos que se utilizan en los problemas y ejercicios están disponibles en línea.
- Se han reorganizado las tablas que se utilizan para el test de signos, el test de rangos con signo y el test de Wilcoxon-Mann-Whitney.

Suplementos para el profesor

Registrándose en la página www.pearsonhighered.com/irc.

Manual en línea de soluciones para el profesor

En este manual se proporcionan las soluciones de todos los ejercicios. Se ha puesto especial atención en asegurar que todos los métodos de solución y notación son consistentes con los utilizados en el texto base. Está disponible para su descarga en el catálogo en línea de Pearson Educación en www.pearsonhighered.com/irc.

Transparencias de Powerpoint

Algunas figuras y tablas seleccionadas del texto están disponibles como transparencias de Powerpoint para su utilización en la creación de presentaciones en Powerpoint para las clases. Se pueden descargar en www.pearsonhighered.com/irc.

Suplementos tecnológicos y opciones de empaquetamiento

Conjuntos de datos

Los conjuntos de datos más grandes utilizados en los problemas y ejercicios del libro se encuentran disponibles como archivos .csv en el sitio web de Pearson de recursos estadísticos y conjuntos de datos:

www.pearsonhighered.com/datasets

StatCrunch™

StatCrunch™ es un software estadístico basado en web que permite a los usuarios realizar análisis complejos, compartir conjuntos de datos y generar informes convincentes basados en los datos analizados. Los usuarios pueden cargar en StatCrunch sus propios datos o buscar en una biblioteca de 12.000 series de datos de dominio público, que cubren prácticamente cualquier tema de interés. Los gráficos interactivos ayudan a los usuarios a entender los conceptos estadísticos, y están disponibles para su exportación, de forma que los informes se puedan enriquecer con representaciones visuales de los datos. Entre las características adicionales se encuentran:

- Un conjunto completo de métodos gráficos y numéricos que permiten a los usuarios analizar y mejorar la interpretación de cualquier conjunto de datos.
- Opciones de informe que permiten a los usuarios crear una amplia variedad de representaciones visualmente impactantes de sus datos.

- Una herramienta de encuestas en línea que permite a los usuarios diseñar y administrar rápidamente encuestas realizadas mediante formularios web.

StatCrunch está disponible para su adquisición cualificada. Para más información, se puede visitar nuestro sitio web www.statcrunch.com.

Hay disponibles también tarjetas de estudio para varias tecnologías, entre las que se incluyen Minitab, SPSS, JMP, StatCrunch, R, Excel y la calculadora gráfica TI.

Agradecimientos de la cuarta edición

La cuarta edición de *Fundamentos de Estadística para las Ciencias de la Vida* mantiene el espíritu y el estilo de redacción de Myra Samuels. Antes de su trágica muerte de cáncer, Myra escribió la primera edición del texto, basándose en su experiencia tanto como profesora de estadística como consultora en estadística. Sin su visión y su esfuerzo no habría existido una primera edición, y desde luego nunca una cuarta.

Muchos investigadores han proporcionado conjuntos de datos al texto, enriqueciéndolo considerablemente. Nos hemos beneficiado de incontables conversaciones a lo largo de los años con David Moore, Dick Scheaffer, Murray Clayton, Alan Agresti, Don Bentley, y muchos otros a los que deseamos dar las gracias.

Deseamos agradecer el entusiasmo y la estupenda labor de guía en los aspectos editoriales de Chris Cummings y Joanne Hill, y la cuidadosa labor de lectura y propuesta de comentarios de Soma Roy. Nuestro agradecimiento también a aquellos que han adoptado la tercera edición y nos han comunicado errores de diverso tipo. En particular, Robert Wolf y Jeff May nos enviaron muchas sugerencias que se han convertido en mejoras de la presente edición. Finalmente, deseamos expresar nuestra gratitud a los revisores de esta edición:

Marjorie E. Bond (Monmouth College), James Grover (University of Texas-Arlington), Leslie Hendrix (University of South Carolina), Yi Huang (University of Maryland, Baltimore County), Lawrence Kamin (Benedictine University), Tiantian Qin (Purdue University), Dimitre Stefanov (University of Akron).

Agradecimientos especiales

A Merrilee, por soportar esas comidas y esas tardes sola mientras yo estaba escribiendo. JAW.

A Michelle y a mis hijos, Ganden y Tashi, por su paciencia y su entusiasmo por este libro. AAS.

INTRODUCCIÓN

Objetivos

En este capítulo presentaremos una serie de ejemplos en las áreas de las Ciencias de la Vida en los que se utiliza la Estadística, con el objeto de entender el alcance de su campo de aplicación. También

- explicaremos en qué se diferencian los experimentos de los estudios observacionales;
- presentaremos los conceptos de efecto placebo, estudio ciego y confusión;
- presentaremos el papel del muestreo aleatorio en estadística.

1.1 La Estadística y las Ciencias de la Vida

Los investigadores en el campo de las Ciencias de la Vida realizan sus actividades en varios entornos: en la clínica, en el laboratorio, en el invernadero, o en el campo. Generalmente, los datos resultantes presentan una cierta *variabilidad*. Por ejemplo, los pacientes que reciben un determinado medicamento responderán de forma algo diferente, los cultivos de células preparados de forma idéntica se desarrollarán de forma algo distinta, o cultivos adyacentes de plantas de trigo genéticamente idénticas producirán cantidades de grano algo diferentes. A menudo, el grado de variabilidad es sustancial, incluso aunque las condiciones experimentales se mantengan tan constantes como sea posible.

El desafío del científico que trabaja en el campo de las Ciencias de la Vida es descubrir los patrones que pueden estar más o menos ocultos por la variabilidad de las respuestas de los sistemas vivos. El científico debe intentar diferenciar la «señal» del «ruido».

La Estadística es la ciencia de la interpretación de los datos y de la toma de decisiones en entornos de variabilidad e incertidumbre. La disciplina de la estadística ha evolucionado en respuesta a las necesidades de los científicos y de aquellos otros cuyos datos presentan variabilidad. Los conceptos y métodos de la estadística permiten al investigador describir la variabilidad y planificar sus investigaciones teniendo en cuenta dicha variabilidad (es decir, para hacer que la «señal» sea fuerte en comparación con el «ruido» de fondo de los datos adquiridos). Los métodos estadísticos se utilizan para analizar los datos y extraer la máxima información, identificando además la fiabilidad de dicha información.

Comenzaremos presentando algunos ejemplos que ilustran el grado de variabilidad que se puede encontrar en datos biológicos y las formas en que dicha variabilidad es un desafío al investigador en biología. Consideraremos brevemente ejemplos que ilustran algunos aspectos estadísticos que aparecen en las investigaciones sobre Ciencias de la Vida e indicaremos dónde se consideran esos aspectos en este libro.

Los primeros dos ejemplos proporcionan un contraste entre un experimento que no presenta variabilidad y otro que presenta una variabilidad considerable.

Ejemplo 1.1.1 Vacuna para el ántrax

El ántrax es una seria enfermedad de las ovejas y el ganado. En 1881, Louis Pasteur realizó un famoso experimento para demostrar el efecto de su vacuna contra el ántrax. Se vacunó contra el ántrax a un grupo de 24 ovejas. Otro grupo de 24 ovejas no vacunadas sirvió como grupo de control. Después, se inoculó a las 48 ovejas un cultivo virulento del bacilo del ántrax. La Tabla 1.1.1 muestra el resultado¹. Los datos de dicha tabla podría decirse que no muestran ninguna variabilidad: todos los animales vacunados sobrevivieron y todos los animales no vacunados murieron. ■

Tabla 1.1.1 Respuesta de las ovejas al ántrax		
Respuesta	Tratamiento	
	Vacunadas	No vacunadas
Mueren de ántrax	0	24
Sobreviven	24	0
Total	24	24
Porcentaje supervivientes	100 %	0 %

Ejemplo 1.1.2 Bacterias y cáncer

Para estudiar el efecto de las bacterias sobre el desarrollo de tumores, los investigadores utilizaron ratones de laboratorio con una alta incidencia natural de tumores de hígado. Un grupo de ratones se mantuvo completamente libre de gérmenes, y el otro se expuso a la bacteria intestinal *Escherichia coli*. La Tabla 1.1.2² muestra la incidencia de tumores de hígado.

Tabla 1.1.2 Incidencia de tumores de hígado en ratones		
Respuesta	Tratamiento	
	<i>E. coli</i>	Sin gérmenes
Con tumores de hígado	8	19
Sin tumores de hígado	5	30
Total	13	49
Porcentaje de tumores de hígado	62 %	39 %

A diferencia de la Tabla 1.1.1, los datos de la Tabla 1.1.2 presentan variabilidad. Los ratones que recibieron el tratamiento no respondieron todos de la misma manera. Debido a esta variabilidad, los resultados de la Tabla 1.1.2 son equívocos. Los datos sugieren que la exposición a *E. coli* aumenta el riesgo de tumores de hígado, pero cabe la posibilidad de que la diferencia de porcentajes (62 % frente a 39 %) pueda ser debida al azar y no realmente a un efecto de la exposición a *E. coli*. Si el experimento se repitiera con diferentes animales, los porcentajes podrían cambiar sustancialmente.

Una forma de explorar qué podría suceder si el experimento se replicara es simular el experimento, lo que se puede realizar como sigue. Tomemos 62 cartas y escribamos «Tumor de hígado» en 27 de ellas ($= 8 + 19$). Escribamos «No tumor de hígado» en las otras 35 ($= 5 + 30$). Barajemos las cartas y escojamos aleatoriamente 13 del mazo (correspondientes a los ratones expuestos a *E. coli*), colocándolas en un montón. Escojamos después otras 49 cartas, también aleatoriamente, y coloquemoslas en un segundo montón. Contemos seguidamente el número de cartas en el «montón de *E. coli*» que tienen escritas las palabras «Tumor de hígado», que corresponden a los ratones expuestos a *E. coli* que han desarrollado tumores de hígado, y veamos si el número es mayor o igual que 8. Este proceso representa la distribución de forma aleatoria de 27 casos de tumores de hígado en dos grupos de ratones (expuestos a *E. coli* y libres de gérmenes), sin que la exposición al *E. coli* haga que sea más ni menos probable que los ratones desarrollen un tumor de hígado, respecto a los ratones libres de gérmenes.

Si repetimos este proceso muchas veces (por ejemplo, 10.000 veces, con la ayuda de un ordenador, en vez de con un mazo físico de cartas), sucede que aproximadamente el 12 % de las veces obtenemos 8 o más ratones expuestos al *E. coli* que desarrollan tumores de hígado. Como algo que sucede el 12 % de las veces no es terriblemente sorprendente, la Tabla 1.1.2 no proporciona una evidencia significativa de que la exposición a *E. coli* aumente la incidencia de tumores de hígado. ■

En el Capítulo 10 trataremos las técnicas estadísticas para la evaluación de los datos como los de las Tablas 1.1.1 y 1.1.2. Por supuesto, en algunos experimentos la variabilidad es mínima y el mensaje que emiten los datos es claro y no precisa ningún tipo de análisis estadístico. Sin embargo, merece la pena destacar que la ausencia de variabilidad es en sí misma un resultado experimental que debe ser justificado con datos suficientes. Por ejemplo, dado que los datos de Pasteur sobre el ántrax (Tabla 1.1.1) no muestran ninguna variabilidad, es verosímil concluir por intuición que los datos proporcionan una prueba «sólida» de la eficacia de la vacuna. Pero, hay que destacar que esta conclusión implica un juicio; consideremos la falta de solidez que tendría la prueba si Pasteur hubiera analizado solo 3 animales en cada grupo, en lugar de 24. Los análisis estadísticos se pueden utilizar para obtener este tipo de juicio, es decir, para determinar si la variabilidad es realmente insignificante. Así pues, una visión estadística resulta útil incluso en la ausencia de variabilidad.

Los dos ejemplos siguientes ilustran las cuestiones adicionales que un enfoque estadístico puede ayudar a resolver.

Ejemplo 1.1.3 Inundación y ATP

En un experimento sobre metabolismo raíz, un fisiólogo de plantas hizo crecer plantas de abedul en semilleros en un invernadero. Inundó con agua cuatro semilleros durante un día y dejó otros cuatro como control. Seguidamente, recolectó los semilleros y analizó la adenosina trifosfato (ATP) de las raíces. Las cantidades medidas de ATP (nmoles por mg de tejido) se muestran en la Tabla 1.1.3 y se representan en la Figura 1.1.1³.

Los datos de la Tabla 1.1.3 plantean varias preguntas. ¿Cómo resumir los valores de ATP de cada condición experimental? ¿Cuánta información proporcionan los datos sobre el efecto de la inundación? ¿Qué confianza se puede tener de que la reducción de ATP en el caso inundado es realmente una respuesta a la inundación y no se debe a una variación aleatoria? ¿Qué tamaño experimental sería necesario para corroborar firmemente el efecto aparente de estos datos? ■

Tabla 1.1.3 Concentración de ATP en raíces de abedul (nmol/mg)	
Inundadas	Control
1,45	1,70
1,19	2,04
1,05	1,49
1,07	1,91

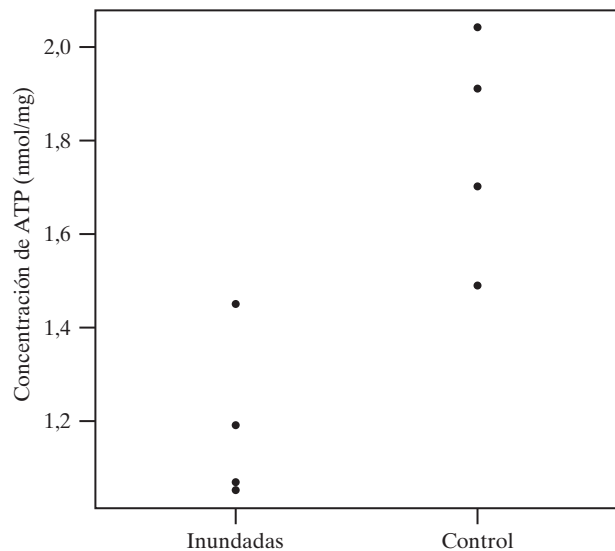


Figura 1.1.1 Concentración de ATP en raíces de abedul

Los Capítulos 2, 6 y 7 consideran preguntas como las planteadas en el Ejemplo 1.1.3. Una cuestión que podemos plantear aquí es si los datos de la Tabla 1.1.3 son consistentes con la afirmación de que la inundación no produce efectos en la concentración de ATP, o bien proporcionan una evidencia significativa de que la inundación afecta a las concentraciones de ATP. Si la afirmación de que no produce efectos es cierta, ¿debería sorprendernos ver que las cuatro observaciones en las que ha existido inundación son menores que las de control? ¿Podría suceder esto solo por casualidad? Si escribiéramos los ocho números 1,05, 1,07, 1,19, 1,45, 1,49, 1,91, 1,70, y 2,04 en tarjetas, las barajáramos y las distribuyéramos aleatoriamente en dos montones, ¿cuál sería la probabilidad de que los cuatro números más pequeños estuvieran en un montón y los cuatro mayores en el otro? Sucede que podemos esperar que ocurra eso 1 vez en 35, por lo que «únicamente el azar» produciría el resultado de la Figura 1.1.1 aproximadamente el 2,9 % de las veces (ya que $1/35 = 0,029$). Por tanto, tenemos alguna evidencia de que la inundación produce algún efecto en la concentración de ATP. Desarrollaremos esta idea con mayor profundidad en el Capítulo 7.

Ejemplo 1.1.4 MAO y esquizofrenia

La monoamina oxidasa (MAO) es una enzima que se piensa tiene un papel en la regulación del comportamiento. Para ver si diferentes categorías de pacientes con esquizofrenia presentan niveles diferentes de actividad MAO, los investigadores recogieron muestras de sangre de 42 pacientes y midieron la actividad MAO en las plaquetas. Los resultados se muestran en la Tabla 1.1.4 y se representan en la Figura 1.1.2 (los valores se expresan en nmol de producto benzilaldehído por 10^8 plaquetas por hora)⁴. Nótese que es mucho más fácil interpretar los datos observando la gráfica (Figura 1.1.2) que leyendo la tabla. El uso de representaciones gráficas es una parte muy importante del análisis de datos. ■

Tabla 1.1.4 Actividad MAO en pacientes con esquizofrenia						
Diagnóstico	Actividad MAO					
I: Esquizofrenia crónica no diferenciada (18 pacientes)	6,8	4,1	7,3	14,2	18,8	
	9,9	7,4	11,9	5,2	7,8	
	7,8	8,7	12,7	14,5	10,7	
	8,4	9,7	10,6			
II: No diferenciada con rasgos paranoides (16 pacientes)	7,8	4,4	11,4	3,1	4,3	
	10,1	1,5	7,4	5,2	10,0	
	3,7	5,5	8,5	7,7	6,8	
	3,1					
III: Esquizofrenia paranoide (8 pacientes)	6,4	10,8	1,1	2,9	4,5	
	5,8	9,4	6,8			

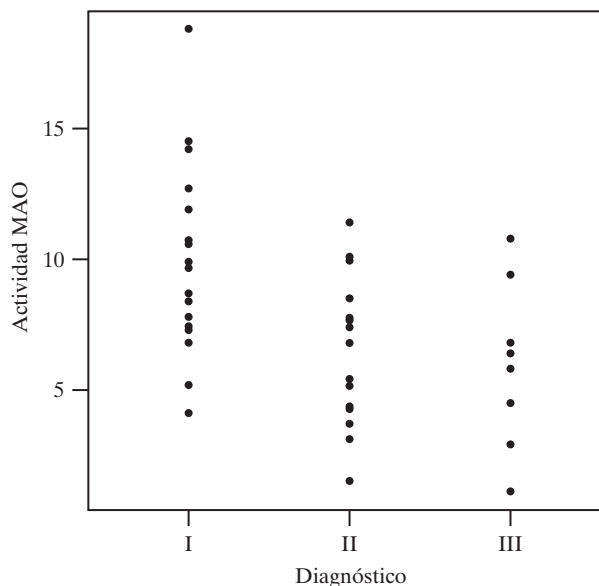


Figura 1.1.2 Actividad MAO en pacientes con esquizofrenia

Para analizar los datos de MAO, lógicamente deben realizarse comparaciones entre los tres grupos de pacientes, para describir la fiabilidad de dichas comparaciones y caracterizar la variabilidad dentro de los grupos. Para ir más allá de los datos y llegar a una interpretación biológica, hay que considerar también aspectos más sutiles, como los siguientes: ¿Cómo se seleccionaron los pacientes? ¿Se escogieron a la vez de la población común de un solo hospital o cada grupo se eligió de un lugar o momento diferente? ¿Se tomó la precaución de que la persona que medía la MAO no supiera el diagnóstico de los pacientes? ¿Consideraron los investigadores diversas formas de subdividir a los pacientes antes de

plantear las categorías diagnósticas concretas de la Tabla 1.1.4? A primera vista, estas cuestiones pueden parecer irrelevantes (¿no deben las medidas hablar por sí mismas?). Veremos, sin embargo, que para una interpretación adecuada de los datos siempre se requiere considerar cuidadosamente la forma en que dichos datos fueron obtenidos.

En los Capítulos 2, 3 y 8 se presentan los aspectos relacionados con la selección de los sujetos experimentales y la forma de evitar el sesgo inconsciente del investigador. En el Capítulo 11 mostraremos cómo escudriñar los datos en busca de patrones puede producir serios errores de interpretación y presentaremos pautas para evitar los riesgos en esas búsquedas.

El siguiente ejemplo muestra cómo los efectos de la variabilidad pueden distorsionar los efectos de un experimento y cómo se puede minimizar dicha distorsión mediante un diseño cuidadoso de dicho experimento.

Ejemplo 1.1.5 Selección de alimento por larvas de insecto

El gorgojo de la raíz del trébol, *Sitona hispidulus*, es una plaga que se alimenta de las raíces de la alfalfa. Una entomóloga realizó un experimento para estudiar la selección de alimento por parte de las larvas de *Sitona*. Su objetivo era investigar qué larvas escogerían preferentemente raíces de alfalfa noduladas (su estado natural) frente a raíces en las que la nodulación hubiera sido suprimida. Las larvas se liberaron en un plato donde tenían disponibles raíces noduladas y no noduladas. Pasadas 24 horas, la investigadora contó las larvas que habían elegido claramente entre los dos tipos de raíces. Los resultados se muestran en la Tabla 1.1.5⁵.

Los datos de la Tabla 1.1.5 parecen sugerir fuertemente que las larvas de *Sitona* prefieren las raíces noduladas. Pero nuestra descripción del experimento ha oscurecido un punto importante. No hemos indicado la disposición de las raíces. Para ver la relevancia de la disposición, supongamos que la experimentadora hubiera utilizado un solo plato, situando todas las raíces noduladas en un lado del mismo y las no noduladas en el otro lado, como se muestra en la Figura 1.1.3(a), y hubiera liberado 120 larvas en el centro del plato. Esta disposición experimental sería seriamente deficiente, ya que entonces los datos de la Tabla 1.1.5 permitirían varias interpretaciones en competencia. Por ejemplo: (a) quizá las larvas prefieren realmente las raíces noduladas, o (b) quizá los dos lados del plato estaban a temperaturas diferentes y las larvas han respondido a la temperatura en vez de a la nodulación, o (c) quizá una larva escogió las raíces noduladas simplemente por casualidad y las otras larvas siguieron su rastro. Dadas estas posibilidades, la disposición experimental que se muestra en la Figura 1.1.3(a) puede resultar en una información débil sobre las preferencias alimentarias de las larvas.

El experimento se dispuso realmente como indica la Figura 1.1.3(b), utilizando seis platos con raíces noduladas y no noduladas dispuestas según un patrón simétrico. Se liberaron veinte larvas en el centro de cada plato. Esta disposición evita los problemas de la disposición de la Figura 1.1.3(a). Dada la alternancia de regiones con raíces noduladas y no noduladas, cualquier fluctuación de las condiciones experimentales (por ejemplo, de la temperatura), tenderá a afectar de la misma forma a los dos tipos. Utilizando varios platos, el experimentador generará datos interpretables incluso aunque las larvas tiendan a seguirse unas a otras. Para analizar apropiadamente el experimento, es necesario conocer el resultado de cada plato. El resumen global de la Tabla 1.1.5 no es adecuado. ■

Tabla 1.1.5 Selección de alimento de las larvas de <i>Sitona</i>	
Selección	Número de larvas
Escogen raíces noduladas	46
Escogen raíces no noduladas	12
Otras (no escogen, mueren, pérdidas)	62
Total	120

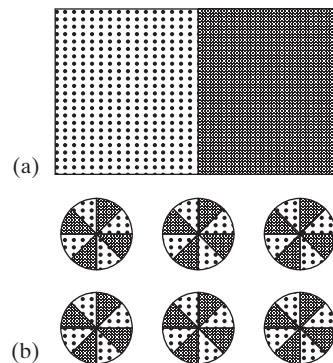


Figura 1.1.3 Posibles disposiciones del experimento de selección de alimento. Las áreas oscuras contienen raíces noduladas y las áreas claras raíces no noduladas. (a) Disposición pobre. (b) Disposición buena

En el Capítulo 11 presentaremos varias formas de disponer los materiales de los experimentos tanto en el espacio como en el tiempo, de forma que se produzcan los resultados que proporcionan mayor información, así como a analizar los datos para extraer la máxima información posible, resistiendo la tentación de sobreinterpretar patrones que pueden representar únicamente una variación aleatoria.

El ejemplo que sigue es un estudio de la relación entre dos cantidades medidas.

Ejemplo 1.1.6 Tamaño corporal y gasto de energía

¿Cuánto alimento necesita una persona? Para investigar la dependencia de los requisitos nutricionales con el tamaño corporal, los investigadores han utilizado técnicas de pesado bajo el agua para determinar la masa corporal libre de grasas de siete personas. Midieron también el gasto de energía en 24 horas en condiciones de actividad tranquila y sedentaria. Las medidas se repitieron dos veces para cada sujeto. Los resultados se muestran en la Tabla 1.1.6 y se representan en la Figura 1.1.4⁶.

Sujeto	Masa libre de grasa (kg)	Gasto de energía en 24 horas (kcal)	
1	49,3	1.851	1.936
2	59,3	2.209	1.891
3	68,3	2.283	2.423
4	48,1	1.885	1.791
5	57,6	1.929	1.967
6	78,1	2.490	2.567
7	76,1	2.484	2.653

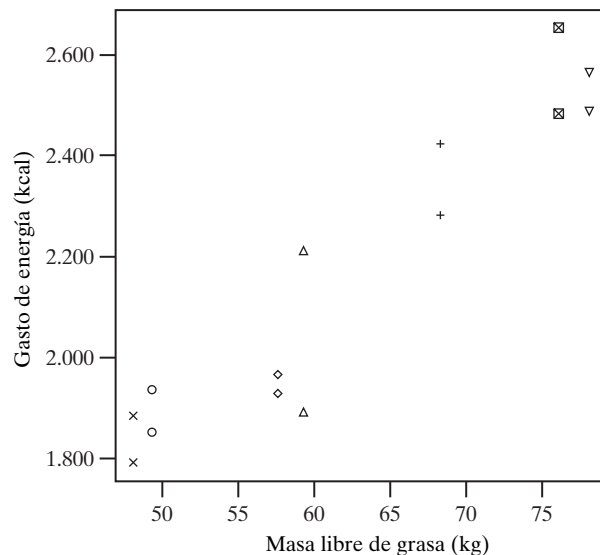


Figura 1.1.4 Masa libre de grasa y gasto de energía de siete personas. Cada persona está representada por un símbolo diferente

Un objetivo primordial en el análisis de estos datos sería describir la relación entre la masa libre de grasa y el gasto de energía, para caracterizar no solo la tendencia general de la relación, sino también el grado de dispersión o de variabilidad de dicha relación. Nótese también que, para analizar los datos, es necesario decidir cómo manejar las observaciones duplicadas de cada sujeto. ■

El centro del Ejemplo 1.1.6 es la relación entre dos variables: masa libre de grasa y gasto de energía. El Capítulo 12 considera métodos para describir estas relaciones, y también para cuantificar la fiabilidad de dichas descripciones.

Una mirada hacia delante

Allí donde es apropiado, los estadísticos hacen uso del computador como herramienta de análisis de datos. A lo largo del libro aparecen resultados y gráficas estadísticas generadas por computador. El computador es una herramienta potente, pero debe ser utilizada con precaución. Al utilizar el computador para realizar cálculos, podemos concentrarnos en los conceptos. El peligro de utilizar el computador en estadística es que se puede pasar directamente a realizar cálculos sin observar cuidadosamente los datos y hacerse las preguntas adecuadas sobre dichos datos. Nuestro objetivo es analizar, comprender e interpretar datos (que son números *en un contexto específico*), no solo realizar cálculos.

Para interpretar adecuadamente un conjunto de datos es necesario conocer cómo y por qué se tomaron. Además de considerar los métodos más ampliamente utilizados en inferencia estadística, consideraremos aspectos relacionados con la toma de datos y el diseño de experimentos. Dichos aspectos, en conjunto, proporcionarán al lector la preparación necesaria para leer la literatura científica y para diseñar y analizar proyectos de investigación sencillos.

Los ejemplos anteriores ilustran el tipo de datos que se van a utilizar en este libro. De hecho, cada uno de ellos volverá a aparecer como ejercicio o ejemplo en un capítulo posterior apropiado. Como muestran los ejemplos, la investigación en las Ciencias de la Vida se ocupa generalmente de la comparación de dos o más grupos de observaciones, o de la relación entre dos o más variables. Comenzaremos nuestro estudio de la estadística con una situación sencilla: observaciones de una *única* variable de un *único* grupo. Muchas de las ideas básicas de la estadística se pueden presentar en este contexto simplificado. Las comparaciones entre dos grupos y los análisis más complicados se presentarán en el Capítulo 7 y en capítulos posteriores.

1.2 Tipos de evidencia

Los investigadores recogen información y realizan inferencias sobre el estado de la naturaleza en muy diversas situaciones. Una buena parte de la estadística trata del *análisis* de datos, pero las consideraciones estadísticas a menudo juegan un papel clave en la planificación y *diseño* de una investigación científica. Comenzaremos presentando ejemplos de las tres clases principales de evidencia que se encuentran a menudo.

Ejemplo 1.2.1 Luz y sordera

El 15 de julio de 1911, la Sra. Jane Decker fue alcanzada por un relámpago en su casa. Sufría de sordera de nacimiento, pero tras el relámpago recuperó la audición, y en el *New York Times* se pudo leer el titular «Un relámpago cura la sordera»⁷. ¿Es esto forzosamente una evidencia de que los relámpagos curan la sordera? ¿Podría haber sido este suceso una mera coincidencia? ¿Existen otras explicaciones para su curación? ■

La evidencia que se presenta en el Ejemplo 1.2.1 es una **evidencia anecdótica**. Una anécdota es una historia corta o un ejemplo de suceso interesante; en este caso, que un relámpago curó una sordera. La acumulación de anécdotas conduce a menudo a conjeturas y a investigación científica, pero son los modelos predecibles, y no las anécdotas, los que establecen una teoría científica.

Ejemplo 1.2.2 Orientación sexual

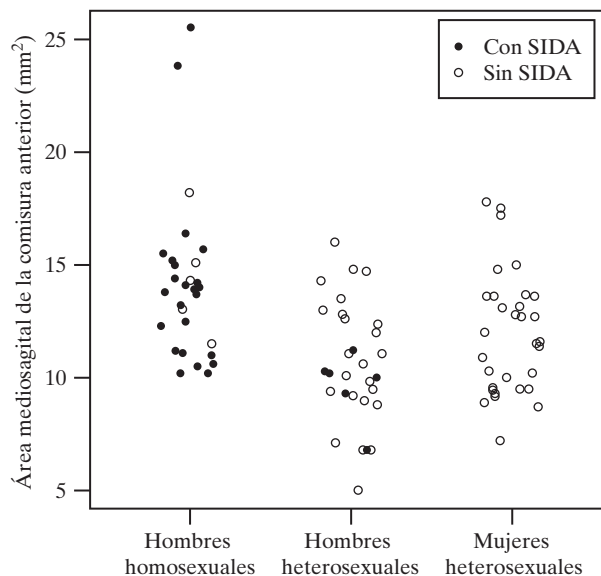
Algunas investigaciones sugieren que existe una base genética para la orientación sexual. En uno de tales estudios se midió el área mediosagital de la comisura anterior (CA) del cerebro de 30 hombres homosexuales, 30 hombres heterosexuales y 30 mujeres heterosexuales. Los investigadores encontraron que la CA tiende a ser mayor en mujeres heterosexuales que en hombres heterosexuales y que era incluso mayor en hombres homosexuales. Estos datos se resumen en la Tabla 1.2.1 y se presentan gráficamente en la Figura 1.2.1.

Los datos sugieren que el tamaño de la CA de los hombres homosexuales es más parecido al de las mujeres heterosexuales que al de los hombres heterosexuales. Al analizar estos datos, hay que tener en cuenta dos cosas: (1) las medidas de dos hombres homosexuales fueron mucho mayores que las del resto; en algunas ocasiones la presencia de uno o dos valores atípicos puede tener un gran impacto en las conclusiones de un estudio. (2) De los 30 hombres homosexuales, 24 padecían SIDA a diferencia del caso de los hombres heterosexuales, solo 6 de 30. Si el SIDA afectara al tamaño de la comisura anterior, entonces este factor podría explicar parte de las diferencias entre los dos grupos de hombres⁸. ■

El Ejemplo 1.2.2 presenta un **estudio observacional**. En este tipo de estudio el investigador recoge de forma sistemática datos de sujetos, pero solo como observador y no como alguien que varía las condiciones. Examinando sistemáticamente todos los datos que surgen de estudios observacionales, es posible evitar los planteamientos selectivos que llevan a informar solo de la evidencia proporcionada por una visión inicial. Sin embargo, los estudios observacionales pueden resultar engañosos debido a las *variables de confusión*. En el Ejemplo 1.2.2 indicamos que el padecer SIDA podría afectar al tamaño de la comisura anterior. Se diría en ese caso que, en este estudio, el efecto del SIDA se confunde con el efecto de la orientación sexual.

Tabla 1.2.1 Área mediosagital de la comisura anterior (mm²)

Grupo	Área mediosagital promedio de la comisura anterior (mm ²)
Hombres homosexuales	14,20
Hombres heterosexuales	10,61
Mujeres heterosexuales	12,03

**Figura 1.2.1** Área mediosagital de la comisura anterior (mm²)

Nótese que el *contexto* en el que surgen los datos es de gran importancia en estadística. Esto resulta bastante claro en el Ejemplo 1.2.2. Los números en sí se pueden utilizar para construir gráficas, como la Figura 1.2.1, pero para entender lo que los datos quieren decir, es necesario tener una comprensión del contexto en el que surgen. El contexto nos indica que hay que estar alerta ante los efectos que otros factores, como el posible impacto del SIDA, puede tener sobre el tamaño de la comisura anterior. Un análisis de datos sin referencia a su contexto carece de sentido.

Ejemplo 1.2.3 Salud y matrimonio

Un estudio realizado en Finlandia obtuvo como resultado que la gente que se casaba con mediana edad tenía una probabilidad menor de desarrollar posteriormente en su vida discapacidades cognitivas (particularmente la enfermedad de Alzheimer)⁹. Sin embargo, mediante un estudio observacional como este no es posible saber si el matrimonio *previene* problemas posteriores, o si por el contrario es menos probable que se casen las personas más propensas a desarrollar problemas cognitivos. ■

Ejemplo 1.2.4 Toxicidad en perros

Antes de probar nuevos medicamentos en pacientes humanos, es práctica común probarlos en perros u otros animales. Como parte de un estudio, se suministró a ocho perros y ocho perras un medicamento en estado de investigación con dosis de 8 mg/kg y 25 mg/kg dentro de cada sexo; las dosis se asignaron de forma aleatoria a los ocho perros y perras. A partir de muestras de sangre se midieron muchos parámetros, como el colesterol, el sodio, la glucosa, etc., para buscar posibles problemas de toxicidad en los perros antes de comenzar los estudios en seres humanos. Uno de los parámetros fue el nivel de fosfatasa alcalina (o NFA, medida en U/l). Los datos se muestran en la Tabla 1.2.2 y se representan en la Figura 1.2.2¹⁰.

El diseño de este experimento posibilita investigar la interacción entre dos factores: el sexo del perro y la dosis. Dichos factores interactuaron de la siguiente forma: en las hembras, el efecto de incrementar la dosis de 8 a 25 mg/kg fue positivo, aunque pequeño (el nivel medio de NFA aumentó de 133,5 a 143 U/l), pero en los machos el efecto de incrementar la dosis de 8 a 25 mg/kg fue negativo (el nivel medio de NFA disminuyó de 143 a 124,4 U/l). En el Capítulo 11 se considerará técnicas para estudiar estas interacciones. ■

Tabla 1.2.2 Nivel de fosfatasa alcalina (U/l)		
Dosis	Macho	Hembra
8	171	150
	154	127
	104	152
	143	105
Promedio	143	133,5
25	80	101
	149	113
	138	161
	131	197
Promedio	124,5	143

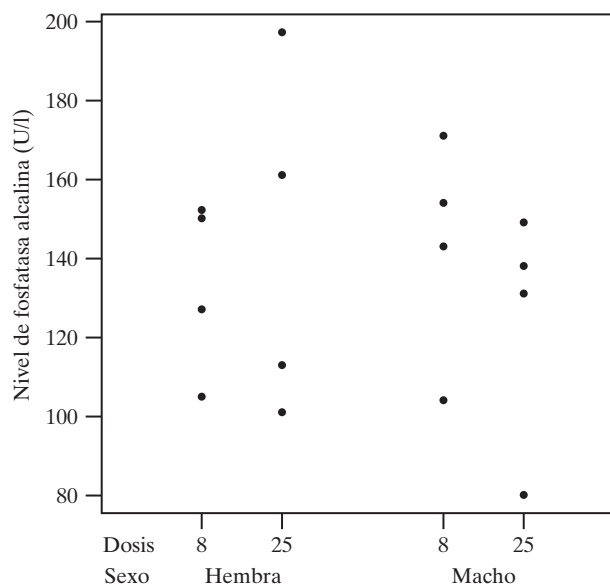


Figura 1.2.2 Nivel de fosfatasa alcalina (U/l)

El Ejemplo 1.2.4 presenta un **experimento**, en el que los investigadores imponen las condiciones (en este caso, las dosis de medicamento) a los sujetos del experimento (los perros). Asignando aleatoriamente los tratamientos (dosis de medicamento) a los sujetos (perros), podemos evitar el problema de la confusión, que complica los estudios observacionales y limita las conclusiones que se pueden obtener de los mismos. Los experimentos aleatorizados se consideran el «estándar de oro» de la investigación científica, pero pueden estar también plagados de dificultades.

Muchas veces, en experimentos con personas se utiliza un **placebo** (una sustancia inerte, como por ejemplo una píldora de azúcar). Es bien conocido que muchas veces las personas presentan una *respuesta al placebo*, es decir, tienden a responder favorablemente a *cualquier* tratamiento, incluso si este es inerte. Este efecto psicológico puede ser muy poderoso. Las investigaciones han demostrado que los placebos pueden ser efectivos para aproximadamente la tercera parte de los enfermos. Es decir, una tercera parte de las personas enfermas indicaron que su enfermedad cesó tras suministrarles un «calmante», es decir, de hecho, una pastilla inerte. En el caso de enfermedades como el asma bronquial, la angina de pecho (dolor recurrente en el pecho causado por un descenso del flujo de sangre al corazón) y úlceras, se ha demostrado que el uso de placebos ha producido resultados clínicos beneficiosos en aproximadamente el 60 % de los pacientes¹¹.

Ejemplo 1.2.5 Autismo

El autismo es una seria enfermedad en la que los niños se apartan de las interacciones sociales normales y algunas veces se sumergen en comportamientos agresivos o repetitivos. En 1997, un niño autista respondió sorprendentemente bien a la hormona digestiva secretina. Se realizó entonces un experimento (una «prueba clínica») en la que se comparó la secretina con un placebo. En este experimento los niños que recibieron secretina mejoraron considerablemente. Sin embargo, los niños que recibieron un placebo mejoraron también considerablemente. No se encontraron diferencias estadísticamente significativas entre los dos grupos. Por tanto, la respuesta favorable del grupo que recibió secretina fue considerada una «respuesta al placebo», y por consiguiente, desafortunadamente, no se pudo concluir que la secretina era beneficiosa (más allá de inducir una respuesta positiva asociada simplemente al hecho de tomar una sustancia como parte de un experimento)¹². ■

La palabra *placebo* significa «me satisfará». La palabra *nocebo* («me perjudicará») se utiliza algunas veces para describir reacciones adversas a riesgos percibidos pero inexistentes. El ejemplo siguiente ilustra la fuerza que pueden tener los efectos psicológicos.

Ejemplo 1.2.6 Asma bronquial

Se dijo a un grupo de pacientes que sufrían asma bronquial que se les iba a suministrar una sustancia química constrictora del pecho. Tras suministrarles dicha sustancia, varios pacientes empezaron a sufrir espasmos bronquiales. Sin embargo, como parte del experimento, se dio a los pacientes una sustancia y se les dijo que aliviaría sus síntomas. En este caso, los espasmos bronquiales se evitaron. En realidad, ambas sustancias eran idénticas: agua destilada. Parece que lo que produjo los espasmos bronquiales fue el poder de la sugestión, el mismo poder que evitó los espasmos¹³. ■

Algo similar al placebo es el tratamiento *falso*, que se puede utilizar en animales y en personas. Un ejemplo de tratamiento falso es inyectar a animales de control una sustancia inerte como suero salino. En algunos estudios sobre tratamientos quirúrgicos, los animales de control (e incluso, en ocasiones, personas) recibe una cirugía «simulada».

Ejemplo 1.2.7 Ligadura de la arteria mamaria

En la década de 1950, la técnica quirúrgica de ligadura de la arteria mamaria interna se convirtió en un tratamiento muy popular para pacientes de angina de pecho. En esta operación el cirujano liga (ata) la arteria mamaria, con el objetivo de aumentar el riego sanguíneo colateral hacia el corazón. Tanto los médicos como los pacientes refrendaron de forma entusiasta la efectividad de este tratamiento. En 1958, se realizaron estudios sobre la ligadura de la arteria mamaria interna en animales, cuyo resultado fue que no era efectiva, lo que produjo dudas sobre su utilidad en humanos. Se realizó entonces un estudio en el que los pacientes se asignaron aleatoriamente a dos grupos. Los pacientes del grupo de tratamiento recibieron la cirugía estándar. A los pacientes del grupo de control se les realizó una operación falsa en la que se realizó la incisión, se puso al descubierto la arteria mamaria como si se tratara de una operación real, pero se cerró la incisión *sin* que la arteria fuera ligada. Los pacientes no tenían forma alguna de saber que su operación había sido falsa. Los porcentajes de mejora en los dos grupos fueron casi idénticos (de hecho, los pacientes que habían sufrido la operación falsa se comportaron ligeramente mejor que los que tuvieron la operación real, pero la diferencia fue pequeña). Un segundo estudio aleatorizado obtuvo también el resultado de que los pacientes que recibieron la cirugía falsa mejoraron igual que los que tuvieron la operación real. Como resultado de estos estudios, los médicos dejaron de utilizar el ligado de la arteria mamaria interna¹⁴. ■

Estudios ciegos

En experimentos con humanos, particularmente aquellos donde se utilizan placebos, se utiliza frecuentemente el concepto de **estudio ciego**. Esto significa que no se revela al sujeto experimental el tratamiento utilizado. El objetivo del estudio ciego es minimizar el alcance de las expectativas del sujeto en la influencia sobre los resultados del experimento. Si los sujetos muestran una reacción psicológica al recibir una medicación, la respuesta al placebo tenderá a equilibrarse entre los dos grupos, por lo que cualquier diferencia entre dichos grupos se podrá atribuir al efecto del tratamiento activo.

En muchos experimentos se ciega también a las personas que estudian las respuestas de los sujetos. Es decir, durante el experimento, dichas personas ignoran la asignación de tratamientos. Consideremos, por ejemplo, lo siguiente:

En un estudio para comparar dos tratamientos para el cáncer de pulmón, un radiólogo examina las placas de rayos X para evaluar el progreso de cada paciente. Las placas de rayos X se codifican de forma que el radiólogo no puede saber qué tratamiento ha recibido cada paciente.

Se alimenta a ratones con tres posibles dietas, y los efectos sobre su hígado son ensayados por un ayudante de investigación que no conoce la dieta que ha recibido cada ratón.

Por supuesto, *alguien* debe saber qué sujeto está en cada grupo, pero esa persona no debe ser la que mida la variable de respuesta. La razón más obvia de cegar a la persona que realiza las evaluaciones es que se reduce la posibilidad de que se produzca un sesgo subjetivo que influya sobre el propio proceso de observación. Alguien que *espera* o *desea* ciertos resultados puede influir inconscientemente en dichos resultados. Este sesgo se puede introducir como medidas aparentemente «objetivas» en forma de variaciones sutiles en las técnicas de disección, procedimientos de valoración, etc.

En estudios médicos sobre seres humanos, el procedimiento de cegado a menudo sirve para propósitos adicionales. Por una parte, muchas veces se pregunta a un determinado paciente si consiente en participar en un estudio médico. Si el

médico que realiza la pregunta ya conoce el tratamiento que va a recibir dicho paciente, entonces el médico podría crear grupos de tratamiento no comparables descartando a ciertos pacientes y aceptando a otros (consciente o inconscientemente). Los efectos de esta asignación sesgada pueden ser sorprendentemente grandes, y se ha percibido que en general favorece al tratamiento «nuevo» o «experimental»¹⁵. Otra razón para realizar estudios médicos ciegos es que el médico puede (de forma consciente o inconsciente) proporcionar un mayor apoyo psicológico, o incluso atender mejor, a aquellos pacientes que están recibiendo el tratamiento que el médico considera mejor.

Un experimento en el que se ciegan tanto los sujetos como las personas que realizan las evaluaciones recibe el nombre de experimento **doble ciego**. El primer experimento sobre la ligadura de la arteria mamaria descrito en el Ejemplo 1.2.7 fue un experimento doble ciego.

Necesidad de grupos de control

Ejemplo 1.2.8 Clofibrate

Se realizó un experimento en el que se suministró a los sujetos el medicamento clofibrate, destinado a bajar el colesterol y por tanto a reducir la probabilidad de muerte por enfermedad coronaria. Los investigadores advirtieron que muchos sujetos no tomaron toda la medicación que el protocolo experimental les indicaba que tenían que tomar. Se calculó el porcentaje de las cápsulas prescritas que había tomado cada sujeto y se dividió a los sujetos en dos grupos de acuerdo al criterio de si habían tomado o no como mínimo el 80 % de las cápsulas que tenían que tomar. La Tabla 1.2.3 muestra que la tasa de mortalidad a los cinco años de aquellos sujetos que tomaron al menos el 80 % de sus cápsulas fue mucho menor que la de los sujetos que no siguieron el protocolo. A primera vista, esto sugiere que tomar la medicación reduce la tasa de mortalidad. Sin embargo, había un grupo de control en el experimento, con un placebo, y muchos sujetos de este grupo de control tomaron menos del 80 % de sus cápsulas. Las tasas de mortalidad de los dos grupos que tomaron el placebo (las de los que siguieron el protocolo y las de los que no lo siguieron) fueron muy similares a las de los dos grupos que tomaron clofibrate.

Tabla 1.2.3 Tasas de mortalidad del experimento con clofibrate				
Clofibrate			Placebo	
Adherencia	<i>n</i>	Mortalidad a 5 años	<i>n</i>	Mortalidad a 5 años
≥ 80 %	708	15,0 %	1.813	15,1 %
< 80 %	357	24,6 %	882	28,2 %

El experimento con clofibrate parece indicar que existen dos clases de sujetos: los que se adhieren al protocolo y los que no. El primer grupo presentó unas tasas de mortalidad mucho menor que el segundo. Esto podría ser debido simplemente a unos mejores hábitos de salud de las personas que son capaces de seguir un protocolo científico durante cinco años que de las personas que no siguen el protocolo. Una conclusión adicional del experimento es que el clofibrate no parece ser más efectivo que el placebo a la hora de reducir la tasa de mortalidad. Si no hubiera existido el grupo de control con el placebo, los investigadores podrían haber obtenido perfectamente una conclusión errónea del estudio y haber atribuido al clofibrate la menor tasa de mortalidad en los que se adhieren al tratamiento, en vez de a otros efectos de confusión que hacen diferentes a las personas que se adhieren al tratamiento de las que no se adhieren¹⁶. ■

Ejemplo 1.2.9 Resfriado común

Hace muchos años, un grupo de investigadores invitó a estudiantes que se suponía particularmente susceptibles al resfriado común a participar en un experimento. Los voluntarios se asignaron aleatoriamente al grupo del tratamiento, en cuyo caso tomaron cápsulas con una vacuna experimental, o al grupo de control, en cuyo caso se les dijo que lo que tomaban era una vacuna, pero de hecho tomaron un placebo (cápsulas que parecían las de la vacuna, pero que contenían lactosa en vez de vacuna)¹⁷. Como se muestra en la Tabla 1.2.4, ambos grupos informaron de una reducción drástica del número de resfriados con respecto a los que habían padecido el año anterior.

Tabla 1.2.4 Número de resfriados en el experimento de vacuna contra el resfriado

	Vacuna	Placebo
<i>n</i>	201	203
Número medio de resfriados		
Año anterior (de memoria)	5,6	5,2
Año en curso	1,7	1,6
% de reducción	70 %	69 %

El número medio de resfriados por persona descendió el 70% en el grupo de tratamiento. Esto podría haber sido una evidencia inicial de que la vacuna produce efecto, si no fuera porque la reducción del grupo de control fue del 69 %.

Se puede atribuir una buena parte del descenso de los resfriados en el Ejemplo 1.2.9 al efecto placebo. Sin embargo, otra preocupación estadística es el **sesgo de panel**, que es un sesgo atribuible a que el propio estudio influye en el comportamiento de los sujetos. Es decir, las personas que saben que están siendo estudiadas cambian su comportamiento. Los estudiantes de este experimento indicaron de memoria el número de catarros que habían tenido el año anterior. El hecho de saberse parte de un estudio podría haber influido en su comportamiento, haciendo menos probable que cogieran un catarro durante el experimento. Participar en el estudio podría haber cambiado la forma en que definían el hecho de tener un resfriado (durante el estudio, se les indicó que informaran al servicio de salud siempre que desarrollaran un resfriado), de modo que, en algunos casos, el resfriado pudo quedar sin informar (¿cómo de enfermo hay que sentirse para decidir que se tiene un resfriado?).

Controles históricos

Los investigadores pueden ser especialmente reacios a utilizar asignaciones aleatorizadas en experimentos con seres humanos. Supongamos, por ejemplo, que los investigadores desean evaluar un nuevo tratamiento prometedor para cierta enfermedad. Podría argumentarse que no es ético negar el tratamiento a ningún paciente, y que por lo tanto todos los pacientes deberían recibir el nuevo tratamiento. Pero entonces, ¿quién serviría como grupo de control? Una posibilidad es utilizar controles históricos, es decir, pacientes anteriores con la misma enfermedad que fueron tratados con otra terapia. Una dificultad de los controles históricos es que a menudo existe una tendencia de los últimos pacientes a mostrar una respuesta mejor (incluso a la misma terapia) que pacientes anteriores con el mismo diagnóstico. Esta tendencia se ha conformado, por ejemplo, comparando a lo largo de los años experimentos realizados en los mismos centros médicos¹⁸. Una razón importante de esta tendencia es que las características globales de la población de pacientes pueden cambiar con el tiempo. Por ejemplo, debido a que las técnicas de diagnóstico tienden a mejorar, pacientes con un diagnóstico determinado (por ejemplo, cáncer de mama) en 2001 pueden tener una mayor esperanza de curación (incluso con el mismo tratamiento) que las que fueron diagnosticadas en 1991, debido a que las primeras han podido ser diagnosticadas con la enfermedad menos avanzada.

Los investigadores médicos no se ponen de acuerdo en la validez y el valor de los controles históricos. El siguiente ejemplo ilustra la importancia de esta controversia.

Ejemplo 1.2.10 Dolencia de la arteria coronaria

La dolencia de las arterias coronarias se trata frecuentemente con cirugía (mediante un *bypass*), pero se puede tratar también únicamente con medicamentos. Muchos estudios han intentado evaluar la efectividad del tratamiento quirúrgico para esta dolencia tan común. En una revisión de 29 estudios de ese tipo, se clasificó cada uno de ellos en función de si utilizaba controles aleatorizados o controles históricos. Las conclusiones de los 29 estudios se muestra en la Tabla 1.2.5¹⁹.

Tabla 1.2.5 Estudios sobre la dolencia de la arteria coronaria			
Tipo de controles	Conclusión sobre la efectividad de la cirugía		Número total de estudios
	Efectiva	No efectiva	
Aleatorizados	1	7	8
Históricos	16	5	21

La Tabla 1.2.5 parece indicar que el entusiasmo por la cirugía es mucho más común entre investigadores que usan controles históricos que entre los que usan controles aleatorizados. ■

Los proponentes del uso de controles históricos argumentan que mediante ajustes estadísticos se pueden realizar comparaciones adecuadas entre un grupo de pacientes actual y un grupo de controles históricos. Por ejemplo, si los pacientes actuales son más jóvenes que los controles históricos, entonces los datos se pueden analizar ajustando, o corrigiendo, el efecto de la edad. Los críticos replican que esos ajustes pueden ser extremadamente inadecuados.

El concepto de controles históricos no se limita a estudios médicos. El tema surge siempre que un investigador compara datos actuales con datos pasados. Procedan los datos del laboratorio, de estudios de campo o de la clínica, el investigador debe afrontar la pregunta: ¿tiene sentido comparar los datos actuales con los pasados? Como mínimo, uno debe siempre preguntarse si el material experimental y/o las condiciones del entorno pueden haberse modificado con el tiempo lo suficiente como para distorsionar la comparación.

Ejercicios 1.2.1-1.2.8

1.2.1 La fluorización del agua potable ha sido siempre un tema controvertido en Estados Unidos. Una de las primeras comunidades que añadió fluoruro en su agua fue Newburgh, Nueva York. En marzo de 1944, se anunció un plan para empezar a añadir fluoruro al agua de Newburgh a partir del 1 de abril de ese año. Durante el mes de abril, los vecinos de Newburgh se quejaron de problemas digestivos, atribuidos a la fluorización del agua. Sin embargo, hubo un retraso en la instalación del equipo de fluorización, de forma que dicha fluorización no empezó realmente hasta el 2 de mayo²⁰. Explique cómo se relaciona el efecto placebo/nocebo con este ejemplo.

1.2.2 La olestra es un aditivo sin calorías ni grasa utilizado en la producción de algunos tipos de patatas fritas. Tras la aprobación del uso de olestra por la Administración de Alimentos y Medicamentos, algunos consumidores se quejaron de que les causaba calambres estomacales y diarrea. Se realizó un experimento doble ciego aleatorizado en el que se dio a algunos sujetos bolsas de patatas fritas con olestra, y a otros, patatas fritas ordinarias. En el grupo de olestra, el 38 % de los sujetos indicaron síntomas gastrointestinales. Sin embargo, en el grupo que tomó patatas fritas normales el porcentaje correspondiente fue del 37 % (los dos porcentajes no son significativamente diferentes, en sentido estadístico)²¹. Explique cómo se relaciona el efecto placebo/nocebo con este ejemplo. Explique también por qué es importante que el experimento sea doble ciego.

1.2.3 (Hipotético) En un estudio sobre acupuntura, se dividen en dos grupos a pacientes con dolores de cabeza. Un grupo recibe acupuntura y el otro toma aspirinas. El acupuntor evalúa la efectividad de la acupuntura y la compara con los resultados del grupo que toma aspirinas. Explique cómo el que el estudio no sea ciego sesga el experimento a favor del acupuntor.

1.2.4 Mediante experimentos aleatorizados controlados se ha encontrado que la vitamina C no es efectiva en el tratamiento de los pacientes con cáncer terminal²². Sin embargo, un artículo de 1976 indicó que los pacientes con cáncer terminal que recibieron vitamina C sobrevivieron mucho más tiempo que los de los controles históricos. Los pacientes tratados con vitamina C fueron seleccionados por cirujanos entre un grupo de pacientes de cáncer de un hospital²³. Explique cómo está sesgado este experimento a favor del uso de la vitamina C.

1.2.5 El 3 de noviembre de 2009, el blog lifehacker.com contenía una entrada de una persona con hongos crónicos en las uñas. Recalcaba que tras muchos años de sufrir e intentar todo tipo de curas, recurrió limarse las uñas dejándolas tan finas como pudo tolerar, y seguidamente aplicó diariamente sobre sus uñas tiritas con vinagre y agua oxigenada. Repitió estas aplicaciones durante 100 días. Tras este tiempo, las uñas crecieron libres de hongos. Utilizando el lenguaje de la estadística, ¿qué clase de evidencia es esta? ¿Es una eviden-

cia convincente de que ese procedimiento es una cura efectiva de los hongos de las uñas?

1.2.6 En cada uno de los siguientes casos [(a), (b) y (c)]

- (I) Indique si el estudio debería ser observacional o experimental.
- (II) Indique si el estudio debería ser ciego, doble ciego o ninguna de las dos cosas. Si el estudio debe ser ciego o doble ciego, explique por qué.
 - (a) Una investigación sobre si el hecho de tomar aspirina reduce las probabilidades de sufrir un ataque al corazón.
 - (b) Una investigación sobre si los bebés nacidos en familias pobres (con renta inferior a 20.000 euros) tienen más probabilidades de pesar menos de 3 kilos cuando nacen que los nacidos en familias adineradas (con rentas superiores a los 65.000 euros).
 - (c) Una investigación sobre si el tamaño del plano mediosagital de la comisura anterior (una parte del cerebro) del hombre está relacionado con su orientación sexual.

1.2.7 (Hipotético) Para evaluar la efectividad de un nuevo fertilizante, los investigadores aplicaron dicho fertilizante a plantas de tomate en el lado oeste de un jardín y no se

aplicó fertilizante a las plantas del lado este de dicho jardín. Posteriormente se pesaron los tomates producidos por cada planta y se encontró que las plantas fertilizadas producían tomates más grandes que los de las plantas no fertilizadas. Se concluyó que el fertilizante funcionaba.

- (a) ¿Se trata de un estudio observacional o experimental? ¿Por qué?
- (b) Este estudio es gravemente defectuoso. Utilice el lenguaje de la estadística para explicar el defecto y cómo afecta a la validez de las conclusiones extraídas por los investigadores.
- (c) ¿Podría haberse utilizado en este caso el concepto de estudio ciego (es decir, se puede aplicar la palabra «ciego» a este estudio? Si es así, ¿cómo?

1.2.8 Los investigadores estudiaron a 1.718 personas mayores de 65 años que vivían en Carolina del Norte. Descubrieron que aquellas que asistían a servicios religiosos de forma regular tenían mayores probabilidades de tener sistemas inmunes fuertes (determinados por los niveles en sangre de la proteína interleukina-6) que los que no asistían regularmente²⁴. ¿Significa esto que asistir a servicios religiosos mejora la salud? ¿Por qué o por qué no?

1.3 Muestreo aleatorio

Para tratar aspectos de la investigación con datos, hay que considerar primero cómo se obtienen dichos datos. La forma de obtención de los datos tiene enormes implicaciones en la selección de los métodos de análisis e incluso en la validez de nuestros estudios. En esta sección examinaremos algunos tipos comunes de métodos de obtención de datos con un énfasis especial en el **muestreo aleatorio simple**.

Muestras y poblaciones

Antes de obtener los datos, debemos considerar primero el alcance de nuestro estudio identificando la **población**. La población está formada por todos los sujetos, animales, especímenes, plantas, etc., de interés. A continuación presentamos unos ejemplos de poblaciones:

- Todas las plantaciones de abedules de Florida.
- Todos los mapaches del Parque Estatal Montaña de Oro.
- Todas las personas con esquizofrenia en Estados Unidos.
- Todos los especímenes de agua de 100 ml de Chorro Creek.

En general, no será posible observar la población completa y, por tanto, deberemos conformarnos con obtener datos de un subconjunto de la población, una **muestra** de tamaño n . A partir de la muestra, realizaremos inferencias sobre la población en conjunto (véase la Figura 1.3.1). Los siguientes ejemplos son de muestras:

- Una selección de 8 ($n = 8$) plantaciones de abedules crecidos en invernadero en Florida.
- Trece ($n = 13$) mapaches capturados con trampas en el camping de Montaña de Oro.
- Cuarenta y dos ($n = 42$) pacientes con esquizofrenia que han respondido a un determinado anuncio de un periódico estadounidense.
- Diez ($n = 10$) muestras de agua de 100 ml obtenidas en un día de 10 poblaciones a lo largo de Chorro Creek.

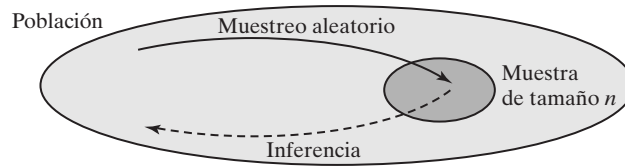


Figura 1.3.1 Muestreo de una población

Observación Existe una confusión potencial entre el significado estadístico del término *muestra* y el sentido en el que se utiliza dicha palabra en biología. Si una bióloga extrae sangre a 20 personas y mide la concentración de glucosa, puede decir que tiene 20 muestras de sangre. Sin embargo, el estadístico dice que tiene *una* muestra de 20 medidas de glucosa: el tamaño de la muestra es $n = 20$. En aras de la claridad, en este libro utilizaremos la palabra *especímen* allí donde un biólogo preferiría utilizar *muestra*. Por tanto, hablaríamos de medidas de glucosa en una muestra de 20 especímenes de sangre.

Idealmente nuestra muestra será un subconjunto representativo de la población. Sin embargo, a menos que procedamos con cuidado, podemos terminar obteniendo una muestra **sesgada**. Una muestra sesgada sobreestima o subestima sistemáticamente una característica de la población. Por ejemplo, consideremos los mapaches capturados con trampas en un camping de la muestra descrita anteriormente. Dichos mapaches podrían diferir sistemáticamente de la población: podrían ser más grandes (por haber tenido un amplio acceso a la comida de los contenedores y de los campistas), menos tímidos (por haber tenido a su alrededor gente que los alimentaba), e incluso podrían haber vivido más tiempo que la población general de mapaches del parque.

Un método que asegura que las muestras serán, a la larga, representativas de la población es el muestreo aleatorio.

Definición de muestreo aleatorio simple

Informalmente, el proceso de obtener una muestra aleatoria simple se puede visualizar mediante papeletas numeradas, como las de las loterías o rifas. Supongamos que cada miembro de la población (es decir, mapache, paciente, planta) se representa con una papeleta que éstas se meten en una caja grande y se mezclan a conciencia. Después, un ayudante con los ojos vendados, extrae n papeletas de la caja y cada vez que se extrae una, las restantes se mezclan de nuevo. Esas n papeletas constituyen la muestra (de forma equivalente, podemos imaginar que hay n ayudantes y que, de forma simultánea, cada uno de ellos toma una papeleta de la caja).

De forma más abstracta, podemos definir el muestreo aleatorio como sigue:

Muestra aleatoria simple

Una **muestra aleatoria simple** de n elementos es una muestra en la que: (a) todos los miembros de la población tienen las mismas oportunidades de ser incluidos en la muestras, y (b) los miembros de la muestra se escogen independientemente entre sí. [El requisito (b) significa que las oportunidades de que un determinado miembro de la población sea elegido no dependen de que otros miembros de la población son elegidos]*.

El muestreo aleatorio simple se puede imaginar de otras formas equivalentes. Podemos imaginar que los miembros de la población se extraen de uno en uno. En el caso de muestreo aleatorio simple, después de cada extracción, todos los

* Técnicamente, el requisito (b) es que cada pareja de miembros de la población tiene las mismas oportunidades de ser seleccionada de la muestra, cada grupo de 3 miembros de la población tiene las mismas oportunidades de ser seleccionada de la muestra, y así sucesivamente. A diferencia de lo anterior, supongamos que tuviéramos una población de 30 personas, y que escribiéramos sus nombres de 3 en 3 en 10 papeletas. Podríamos entonces seleccionar una papeleta para obtener una muestra de tamaño $n = 3$, pero esto no sería un muestreo aleatorio simple, ya que la pareja (1, 2) podría estar en la muestra pero a lo mejor la pareja (1, 4) no. En este caso las extracciones de los miembros de la muestra no son independientes entre sí (esta clase de muestreo se conoce como «muestreo por grupos», con 10 grupos de tamaño 3). Si la población es infinita, entonces la definición técnica de que todos los subconjuntos de un determinado tamaño tienen las mismas oportunidades de ser elegidos como parte de la muestra es equivalente al requisito de que todos los miembros de la muestra son elegidos independientemente.

miembros restantes de la población tienen las mismas oportunidades de ser elegidos. Otra forma de imaginarlo es considerar la totalidad de las muestras posibles de tamaño n . Si todas las muestras posibles tienen las mismas oportunidades de ser obtenidas, el proceso resultante es un muestreo aleatorio simple.

Uso de la aleatoriedad

Cuando se realizan investigaciones estadísticas, es necesario utilizar la aleatoriedad. Como se ha comentado anteriormente, la aleatoriedad se utiliza en el muestreo aleatorio simple (todos los miembros de la población tienen las mismas posibilidades de ser seleccionados). En el Capítulo 7 comentaremos experimentos en los que se desea comparar los efectos de diferentes tratamientos en los miembros de una muestra. Para realizar esos experimentos deberemos asignar aleatoriamente los tratamientos a los sujetos (de forma que cada sujeto tenga las mismas posibilidades de recibir el tratamiento A o el tratamiento B).

Desafortunadamente, en términos prácticos, los seres humanos no somos muy capaces de utilizar mentalmente la aleatoriedad. No somos capaces de eliminar el sesgo inconsciente que a menudo conduce sistemáticamente a incluir o excluir a ciertos individuos en nuestra muestra (o al menos a aumentar o disminuir las posibilidades de escoger a ciertos individuos). Por este motivo, cuando deseamos escoger individuos en un muestreo aleatorio, debemos utilizar recursos externos en forma de elementos mecánicos dados, monedas o billetes de lotería, dispositivos electrónicos que producen dígitos aleatorios, como computadores y calculadoras o tablas de números aleatorios, como la Tabla 1 que se incluye al final del libro. Aunque sencillos, los métodos mecánicos como la colocación de papeletas en una caja no son prácticos, por lo que nos centraremos en el uso de dígitos aleatorios para la selección de muestras.

Cómo escoger una muestra aleatoria

A continuación se presenta un procedimiento simple para escoger una muestra aleatoria de n elementos a partir de una población con un número de elementos finito.

- Crear una **estructura de muestreo**: una lista con todos los elementos de la población identificados cada uno de ellos por un número único. Todos los números de identificación deben tener el mismo número de dígitos. Por ejemplo, si la población tiene 75 elementos, los números de identificación podrían ser 01, 02, ..., 75.
- Leer números de la Tabla 1, de una calculadora o de un computador. Rechazar los números que no correspondan a ningún miembro de la población (por ejemplo, si la población tiene 75 elementos a los que hemos asignado los números de identificación 01, 02, ..., 75, no se considerarán los números 76, 77, ..., 99 y 00). Continuar hasta que se hayan generado n números (ignórese cualquier aparición repetida del mismo número).
- Los miembros de la población con los números de identificación seleccionados constituyen la muestra.

El ejemplo siguiente ilustra este procedimiento.

Ejemplo 1.3.1

Supongamos que deseamos escoger una muestra aleatoria de tamaño 6 de una población con 75 miembros. Etiquetamos los miembros de la población con los números 01, 02, ..., 75. Utilizando la Tabla 1 u otro procedimiento generamos una secuencia de números aleatorios*. Por ejemplo, una calculadora podría producir la siguiente secuencia:

838717940162534597539822

Al examinar las parejas de dos números, ignoraremos los mayores que 75, así como aquellas parejas que ya se hayan elegido previamente.

83 87 17 94 01 62 53 45 97 53 98 22

Por tanto, la muestra estará formada por los miembros de la población con los siguientes números de identificación: 17, 01, 62, 53, 45, 22. ■

* La mayoría de las calculadoras generan números aleatorios expresados como números decimales entre 0 y 1; para convertirlos en dígitos aleatorios, basta con ignorar el cero inicial y leer los dígitos que siguen a la coma decimal. Para generar una secuencia larga de dígitos aleatorios basta con llamar repetidamente a la función de generación de números aleatorios de la calculadora.

Observación Al llamar *aleatorios* a los dígitos que se obtienen de la Tabla 1, de la calculadora o computador, utilizamos el término *aleatorio* de forma un tanto flexible. Estrictamente hablando, los dígitos aleatorios son los producidos por un *proceso* aleatorio (por ejemplo, lanzar un dado con 10 caras). Los dígitos que se obtienen de la Tabla 1, la calculadora o computador son realmente dígitos *seudoaleatorios*. Son generados por un proceso determinista (aunque posiblemente muy complejo), diseñado para producir secuencias de dígitos que imitan a las generadas de forma verdaderamente aleatoria.

Observación Si la población es grande, entonces el *software* de los computadores puede resultar de gran ayuda para generar la muestra. Si se necesita una muestra aleatoria de tamaño 15 de una población con 2.500 miembros, se utilizará el computador (o la calculadora) para generar 15 números aleatorios entre 1 y 2.500 (si se producen duplicados en el conjunto de números, basta con volver atrás y generar más).

Aspectos prácticos del muestreo aleatorio

En muchos casos, la obtención de una muestra aleatoria apropiada resulta difícil o imposible. Por ejemplo, para obtener una muestra aleatoria de los mapaches del Parque Estatal Montaña de Oro sería necesario crear primero la estructura de muestreo, proporcionando a cada mapache del parque un único número. Después, tras generar la lista de números aleatorios que identifica nuestra muestra, sería necesario capturar a los mapaches cuyo número coincide con el de la lista. Es una tarea prácticamente imposible.

En la práctica, si es posible obtener una muestra aleatoria, debe hacerse. Cuando no es práctico obtenerla, es importante tomar precauciones para asegurar que los sujetos del estudio se pueden ver *como si* se hubieran obtenido mediante el muestreo aleatorio de una población. Es decir, la muestra debe estar formada por individuos con las mismas oportunidades de ser escogidos de la población, y los individuos se deben escoger independientemente. Para conseguir esto, el primer paso es definir la población. El siguiente paso es examinar el procedimiento de selección de las unidades observacionales y preguntarse: ¿podrían estas *observaciones* haber sido seleccionadas aleatoriamente? En el ejemplo de los mapaches, esto quiere decir que en primer lugar debería definirse la población de mapaches creando una frontera geográfica definida basándose en el hábitat de los mapaches, y después situar trampas en lugares escogidos aleatoriamente dentro del hábitat de la población, usando diferentes tipos de cebo y tamaño de trampa (podríamos utilizar números aleatorios para generar coordenadas de latitud y longitud dentro del hábitat de la población). Aunque lejos de ser ideal (algunos mapaches pueden ser huidizos de las trampas y los cachorros pueden no entrar en ninguna trampa), esto es ciertamente mejor que simplemente capturar mapaches en un lugar atípico del parque escogido de forma no aleatoria (por ejemplo, en un camping). Es de suponer que ahora una gran mayoría de los mapaches tiene las mismas posibilidades de ser capturados (es decir, las mismas oportunidades de ser seleccionados), y además la captura de un mapache no tiene ninguna influencia en la captura de otro (es decir, se pueden considerar escogidos de forma independiente). Entonces, es razonable tratar las observaciones como si se hubieran escogido aleatoriamente.

Métodos de muestreo aleatorio no simple

Existen otras clases de muestreo que son aleatorios en algún sentido, pero que no son simples. Dos técnicas comunes de muestreo aleatorio no simple son el **muestreo aleatorio por grupos** y el **muestreo aleatorio estratificado**. Para ilustrar el concepto de muestreo por grupos, consideremos una modificación del método de la lotería para generar una muestra aleatoria simple. En el caso de muestreo por grupos, en vez de asignar una única etiqueta (o número identificativo) a cada miembro de la población, se asignan identificadores a grupos enteros de individuos. A medida que se extraen papeletas de la caja, se seleccionan grupos enteros de individuos, como se muestra en el siguiente ejemplo y en la Figura 1.3.2.

Ejemplo 1.3.2 El cardo La Graciosa

El cardo La Graciosa (*Cirsium loncholepis*) es una planta nativa de las Dunas de Gaudalupe en la costa central de California, que se encuentra en peligro de extinción. En un estudio sobre germinación de semillas, se escogieron aleatoriamente 30 plantas de la población de plantas de las Dunas de Guadalupe y se recogieron todas las semillas de esas 30 plantas. Las semillas forman una muestra por grupos de la población de todas las semillas del cardo La Graciosa en Guadalupe, donde cada planta identifica a uno de los grupos²⁵. ■

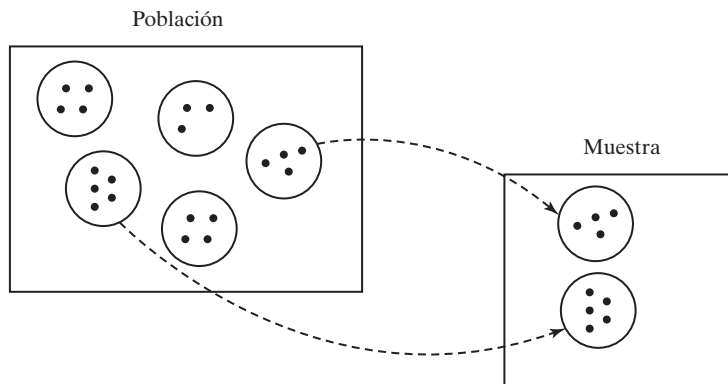


Figura 1.3.2 Muestreo aleatorio por grupos. Los puntos representan individuos de la población que se juntan en grupos (círculos). Para obtener la muestra, los individuos de la población se muestrean por grupos

Un muestreo aleatorio estratificado se realiza dividiendo previamente a la población en **estratos** (conjuntos homogéneos de individuos). Seguidamente, se toman muchas muestras aleatorias simples (una por cada estrato) y se combinan para formar la muestra completa (véase la Figura 1.3.3). A continuación se presenta un ejemplo de muestreo aleatorio estratificado.

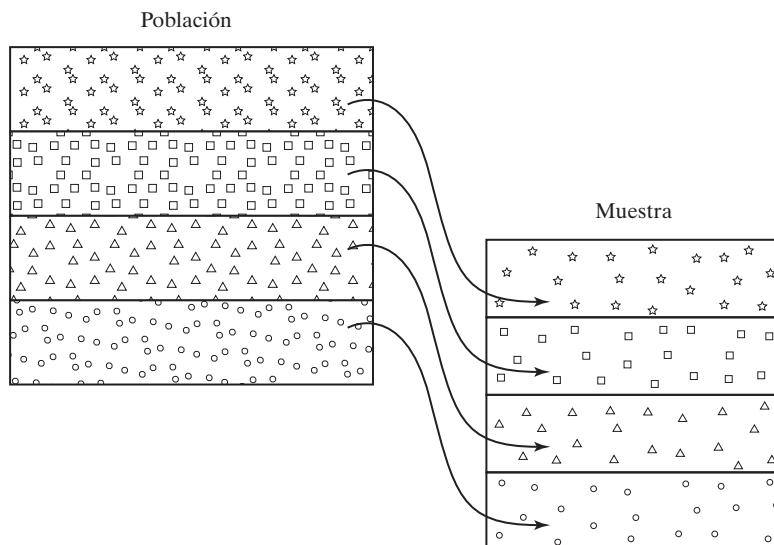


Figura 1.3.3 Muestreo aleatorio estratificado. Los puntos representan individuos de la población que se agrupan en estratos. Los individuos de cada estrato se muestrean aleatoriamente y se combinan para formar la muestra

Ejemplo 1.3.3 Cangrejos de la arena

En un estudio sobre el parasitismo de los cangrejos de la arena (*Emerita analoga*), los investigadores realizaron un muestreo aleatorio estratificado dividiendo una playa en bandas de 5 metros de anchura paralelas a la orilla. Dichas bandas se utilizaron como estratos debido a que la carga de parásitos de los cangrejos puede diferir sistemáticamente en función de la distancia a la orilla, con lo que la carga de parásitos dentro de cada estrato es más similar que las cargas entre estratos diferentes. El primer estrato era la banda de 5 metros justo debajo de la superficie del agua paralela a la orilla. El segundo estrato era la banda de 5 metros fuera del agua y pegada a la orilla, y después,

alejándose de la orilla, las bandas de 5 metros correspondientes al tercer y cuarto estrato. En cada estrato se muestrearon aleatoriamente 25 cangrejos, por lo que el tamaño total de la muestra fue de 100 cangrejos²⁶. ■

La mayoría de los métodos estadísticos presentados en este texto suponen que estamos trabajando con datos procedentes de un muestreo aleatorio simple. Las muestras generadas mediante muestreo aleatorio simple se denominan generalmente *muestras aleatorias*. Pero nótese que es realmente el *proceso* de muestreo y no la propia muestra lo que se define como aleatorio. La aleatoriedad no es una propiedad de una muestra concreta, una vez que ya ha sido elegida.

Error de muestreo

¿Cómo podemos dar una base lógica para el hecho de que la inferencia a partir de una muestra limitada se pueda aplicar a una población mucho mayor? El planteamiento de la teoría estadística es utilizar un modelo idealizado de la relación muestra-población. En este modelo, denominado **modelo de muestreo aleatorio**, la muestra de la población se escoge mediante dicho muestreo aleatorio: El modelo se representa esquemáticamente en la Figura 1.3.1.

El modelo de muestreo aleatorio es útil porque proporciona la base para responder a una pregunta: ¿Cómo de probable es que la muestra sea representativa de la población? El modelo se puede utilizar para determinar en qué medida una inferencia puede estar influenciada por el azar, o ser «cuestión de suerte». De forma más explícita, una muestra elegida aleatoriamente en general no se asemejará exactamente a la población de la que ha sido extraída. Esta discrepancia entre la muestra y la población se denomina **error de muestreo debido al azar**, o **error de muestreo**. En capítulos posteriores veremos cómo la teoría estadística que se obtiene a partir del concepto de muestreo aleatorio permite poner límites en la probabilidad del error debido al muestreo en un experimento. La cuantificación de este error es una de las principales contribuciones de la teoría estadística al pensamiento científico.

Como nuestras muestras se escogen aleatoriamente, siempre estará presente el error de muestreo. Sin embargo, si se muestrea de forma no aleatoria, el efecto de este error se puede agravar de forma impredecible introduciendo **sesgo de muestreo**, que es una tendencia sistemática que presentan ciertos elementos de la población a ser seleccionados de forma más probable que otros. Los dos ejemplos siguientes ilustran el sesgo de muestreo.

Ejemplo 1.3.4 Longitudes de peces

Un biólogo planea estudiar la distribución de la longitud del cuerpo de una cierta población de peces en la Bahía Chesapeake. La muestra se recoge utilizando una red de pesca. Será entonces más probable que los peces pequeños escapen por los agujeros de la red. Por tanto, será menos probable recoger peces pequeños que grandes, por lo que el procedimiento de muestreo está sesgado. ■

Ejemplo 1.3.5 Tamaños de células nerviosas

Un neuroanatomista planea medir los tamaños de células nerviosas en el tejido cerebral de gato. Al examinar el espécimen de tejido, el investigador debe decidir cuáles de los cientos de células presentes se seleccionarán para ser medidas. Algunas células nerviosas estarán incompletas debido al corte con el microtomo realizado para seleccionar el tejido. Si la medida del tamaño se realiza solo sobre células completas, aparece un sesgo debido a que las células pequeñas tienen una mayor probabilidad de no ser afectadas por la cuchilla del microtomo. ■

Cuando el procedimiento de muestreo está sesgado, la muestra puede no representar adecuadamente a la población, debido a que se encuentra sistemáticamente distorsionada. Por ejemplo, en el Ejemplo 1.3.4, los peces pequeños tienden a estar menos representados en la muestra, de forma que la longitud de los peces en la muestra tenderá a ser mayor que en la población.

El ejemplo siguiente ilustra un tipo de no aleatoriedad diferente del sesgo.

Ejemplo 1.3.6 Sacarosa en raíces de remolacha

Una agrónomo planea muestrear raíces de remolacha en un campo de cultivo para medir su contenido de sacarosa. Supongamos que toma todos sus especímenes de una zona pequeña del campo de cultivo seleccionada aleatoriamente. El procedimiento de muestreo no estaría sesgado pero tendería a producir una muestra *demasiado homogénea*, debido a que la muestra no reflejaría las variaciones ambientales y de entorno a lo largo de todo el campo de cultivo. ■

El Ejemplo 1.3.6 ilustra un principio importante que se pasa a veces por alto en el análisis de los datos. Para comprobar la aplicabilidad del modelo de muestreo aleatorio, es necesario preguntar no solo si el procedimiento de muestreo

puede estar sesgado, sino también si el procedimiento de muestreo refleja adecuadamente la variabilidad inherente a la población. Información falsa sobre esa variabilidad puede distorsionar las conclusiones científicas de forma tan seria como el sesgo.

Consideraremos seguidamente algunos ejemplos donde se puede aplicar razonablemente el modelo de muestreo aleatorio.

Ejemplo 1.3.7 Resistencia del maíz a los hongos

Una cierta variedad de maíz es resistente a la infección por hongos. Para estudiar la herencia de esta resistencia, un agrónomo cruzó la variedad resistente con una variedad no resistente y midió la resistencia en las plantas de la descendencia. La descendencia real del experimento se puede ver como una muestra aleatoria de una población conceptual de toda la descendencia *potencial* de un cruce particular. ■

Cuando el objetivo de un estudio es *comparar* dos o más condiciones experimentales, una definición muy estrecha de la población puede resultar ventajosa, como se ilustra en el ejemplo siguiente.

Ejemplo 1.3.8 Metabolismo de nitritos

Para estudiar la conversión de nitrito en nitrato en la sangre, los investigadores inyectaron a cuatro conejos blancos de Nueva Zelanda una solución de moléculas de nitrito marcadas radiactivamente. Diez minutos después de la inyección, midieron en cada conejo el porcentaje de nitrito que se había transformado en nitrato²⁷. Aunque los cuatro animales no fueron escogidos de la población de forma literalmente aleatoria, podría no obstante ser razonable asumir que las medidas del metabolismo del nitrito son una muestra aleatoria de medidas similares en conejos blancos de Nueva Zelanda (asumiendo que la edad y el género son irrelevantes para el metabolismo del nitrito). ■

Ejemplo 1.3.9 Tratamiento de la colitis ulcerosa

Un equipo médico realizó un estudio de dos terapias, A y B, para el tratamiento de la colitis ulcerosa. Todos los pacientes del estudio eran derivados a una clínica de una gran ciudad. Cada paciente se observó en busca de una «respuesta» satisfactoria a la terapia. Aplicando el modelo de muestreo aleatorio, el objetivo de los investigadores sería realizar inferencias sobre la población de todos los pacientes de colitis ulcerosa derivados a clínicas urbanas. Consideremos primero la inferencia sobre las posibilidades reales de respuesta al tratamiento. Dicha inferencia sería válida si la probabilidad de respuesta de cada terapia es la misma en todas las clínicas urbanas con pacientes derivados de este tipo. Sin embargo, este supuesto podría ser algo cuestionable, haciendo que los investigadores desearan que la población se definiera de forma más restringida (por ejemplo, como «el tipo de pacientes con colitis ulcerosa derivados a una determinada clínica»). Incluso una población tan restringida podría resultar de interés en un estudio comparativo. Por ejemplo, si el tratamiento A es mejor que el tratamiento B en la población restringida, podría ser razonable inferir que A sería mejor que B en una población más amplia (incluso aunque las probabilidades de respuesta real pudieran ser diferentes en la población más amplia). De hecho, incluso podría argumentarse que una población más amplia podría incluir todos los pacientes de colitis ulderosas, no solo los derivados a clínicas urbanas. ■

En la investigación sucede a menudo que, por razones prácticas, la población que se estudia realmente es más limitada que la población real de interés. Para aplicar el tipo de razonamiento que se ilustra en el Ejemplo 1.3.9, hay que justificar que los resultados en la población restringida (o, al menos algunos aspectos de sus resultados), se pueden extrapolar con sentido a la población de interés. Esta extrapolación no es una inferencia *estadística*, y debe defenderse en el terreno biológico, no estadístico.

En la Sección 2.8 trataremos la conexión entre muestras y poblaciones a medida que desarrollamos el concepto de inferencia estadística.

Errores no de muestreo

Además de los errores de muestreo, en los estudios estadísticos pueden surgir otros motivos de preocupación. Un **error no de muestreo** es un error que no está causado por el método de muestreo, es decir, que podría haber aparecido incluso aunque el investigador tuviera un censo de toda la población. Por ejemplo, la forma en que se formulan las preguntas puede influir grandemente en la manera de responderlas de la gente, como muestra el Ejemplo 1.3.10.

Ejemplo 1.3.10 Ayudas sobre el aborto

En 1991, el Tribunal Supremo de Estados Unidos dictó una polémica sentencia prohibiendo la orientación sobre el aborto en las clínicas de planificación familiar financiadas con fondos federales. Poco tiempo después de la sentencia, se preguntó a una muestra de 1.000 personas: «Como usted sabe, el Tribunal Supremo de Estados Unidos sentenció recientemente que el gobierno federal no está obligado a usar fondos provenientes de los impuestos para realizar, aconsejar o remitir al aborto como método de planificación familiar. En general, ¿está usted a favor o en contra de esta sentencia?». En la muestra, el 48 % estuvo a favor, el 48 % en contra y el 4 % no tuvo opinión.

En un sondeo diferente realizado casi al mismo tiempo, pero con una organización distinta, se preguntó a 1.200 personas: «¿Está usted a favor o en contra de que la decisión del Tribunal Supremo impida que los médicos y personal sanitario de las clínicas hable del aborto en clínicas de planificación familiar que reciben fondos federales?». En esta muestra, el 33 % estuvo a favor y el 65 % en contra²⁸. La diferencia en los porcentajes a favor de la opinión es demasiado grande como para atribuirlos a errores en el muestreo debidos al azar. Parece que la forma de formular la pregunta tiene un fuerte impacto en las respuestas. ■

Otro tipo de error no de muestreo es el **sesgo de no respuesta**, que es el sesgo causado por personas que no responden a alguna de las preguntas o no devuelven una encuesta cuando esta es por escrito. Es habitual que solo la tercera parte de los que reciben una encuesta por correo se la devuelvan rellena a los investigadores (consideramos las personas que reciben la encuesta como parte de la muestra, tanto si no la completan del todo como si no la completan en absoluto). Si las personas que responden son distintas de las que no lo hacen (y este es a menudo el caso, ya que la gente con relaciones fuertes con algún tema tiende a rellenar un cuestionario sobre el mismo, mientras que el resto tenderá a ignorarlo), entonces los datos recogidos no representarán de forma precisa a la población.

Ejemplo 1.3.11 Test del VIH

Se preguntó a una muestra de 949 hombres si se someterían a un test de VIH en su sangre. De los 782 que aceptaron realizar el test, se encontró que 8 (1,02 %) eran seropositivos. Sin embargo, algunos hombres no quisieron realizar el test. Los investigadores en salud que realizaron el estudio tuvieron acceso a algunos especímenes de suero que se habían tomado previamente de los 167 hombres que habían rechazado hacer la prueba, y encontraron que 9 de ellas (5,4 %) eran seropositivas²⁹. Por tanto, era mucho más probable que aquellos que rechazaron realizar la prueba fueran seropositivos. Una estimación del porcentaje de seropositivos basada únicamente en las personas que aceptaron realizar la prueba es muy probable que subestimara la preponderancia real. ■

Existen otros casos en los que el experimento se enfrenta con el irritante problema de **pérdida de datos**, es decir, observaciones que estaban planeadas pero que no se realizaron. Además de la falta de respuestas, esto puede producirse por muerte de los animales o de las plantas experimentales, debido a fallos en los equipos, o debido a que los sujetos humanos no dan respuesta en observaciones de seguimiento.

Un enfoque habitual a este problema de la pérdida de datos es utilizar simplemente los datos restantes e ignorar que algunas observaciones se han perdido. Este enfoque es tentadoramente simple, pero debe utilizarse con extremo cuidado, ya que las comparaciones basadas en los datos restantes pueden estar seriamente sesgadas. Por ejemplo, si se pierden observaciones de algunos ratones experimentales debido a que dichos ratones mueren por causas debidas al tratamiento que han recibido, obviamente no es válido comparar solo a los ratones que han sobrevivido. Como otro ejemplo, si los pacientes abandonan un estudio médico porque piensan que su tratamiento no está funcionando, entonces los análisis de los restantes pacientes podrían dibujar una situación enormemente distorsionada.

Naturalmente, es mejor realizar todos los esfuerzos posibles para evitar la pérdida de datos. Pero si se pierden datos, es crucial que las posibles razones de esa pérdida se consideren en la interpretación y el informe de resultados.

Los datos pueden ser también engañosos si existe un sesgo en la forma de obtenerlos. Las personas tienen dificultades en recordar las fechas en las que suceden eventos y tienden a dar respuestas poco fiables incluso si se les hacen preguntas como «¿Cuántas veces por semana hace usted ejercicio?». También pueden estar sesgadas al realizar observaciones, como muestra el siguiente ejemplo.

Ejemplo 1.3.12 Azúcar e hiperactividad

Un conjunto de madres que pensaban que sus hijos pequeños eran «sensibles al azúcar» se dividió aleatoriamente en dos grupos. A las del primer grupo se les dijo que se había dado a sus hijos una dosis grande de azúcar, mientras que a las del segundo grupo se les dijo que se había dado a sus hijos un placebo. De hecho, a todos los niños se les dio un placebo. No obstante, las madres del primer grupo indicaron que sus hijos habían sido mucho más hiperactivos durante el periodo de estudio de 25 minutos que las madres del segundo grupo³⁰. Se realizaron medidas neutrales que indicaron que los niños del primer grupo fueron realmente un poco *menos* activos que los del segundo grupo. Muchos otros estudios realizados no habían podido encontrar una relación entre el consumo de azúcar y la actividad de los niños, a pesar de la creencia muy extendida de que el azúcar produce un comportamiento hiperactivo. Parece que las expectativas de esas madres habían influido en sus observaciones³¹. ■

Ejercicios 1.3.1-1.3.6

1.3.1 En cada uno de los estudios siguientes, identifique qué técnica de muestreo describe mejor la forma en que se tomaron los datos (o la forma en que se podrían haber tomado si se tomaran): muestreo aleatorio simple, muestreo aleatorio por grupos o muestreo aleatorio estratificado. En los casos de muestreo por grupos, identifique los grupos, y en los casos de muestreo estratificado, los estratos.

- Todos los 257 pacientes de leucemia de tres clínicas pediátricas elegidas aleatoriamente en Estados Unidos fueron inscritos para participar en una prueba clínica de un nuevo medicamento.
- Un total de doce especímenes de 10 g de suelo se recogieron en sitios aleatorios de una granja para estudiar las propiedades físicas y químicas del suelo.
- En un estudio sobre contaminación se recogieron tres especímenes de 100 ml de aire en cuatro altitudes diferentes (100 m, 500 m, 1.000 m y 2.000 m), es decir, un total de doce especímenes.
- Un total de 20 uvas se recogieron de vides aleatorias en un viñedo para evaluar si se podía proceder a la vendimia.
- Veinticuatro perros (ocho escogidos aleatoriamente de razas pequeñas, ocho escogidos aleatoriamente de razas de tamaño medio y ocho escogidos aleatoriamente de razas de tamaño grande) se escogieron para participar en un experimento de evaluación de un nuevo programa de entrenamiento.

1.3.2 En cada uno de los estudios siguientes, identifique la fuente o fuentes de sesgo en el muestreo y describa: (i) cómo podría afectar a las conclusiones del estudio y (ii) cómo se podría modificar el método de muestreo para evitar el sesgo.

- Se reclutaron 800 voluntarios en clubes nocturnos para participar en un experimento para evaluar un nuevo tratamiento de la ansiedad social.
- En un estudio sobre contaminación del agua, se recogieron especímenes de agua de un arroyo en 15 días lluviosos.
- Para estudiar la distribución del tamaño (radio) de arbustos de roble (árboles de roble tipo arbusto), se seleccionaron 20 robles utilizando coordenadas de longitud y

latitud aleatorias. Si la coordenada aleatoria caía dentro de la copa del árbol, dicho árbol era seleccionado, si no, se generaba otra coordenada aleatoria.

- Para estudiar la distribución de tamaños de los bacalao de roca (*Epinephelus puscus*) lejos de la costa del sudeste de Australia, se anotaron las longitudes y los pesos de todos los bacalao capturados por un barco de pesca comercial en un día (utilizando métodos tradicionales de anzuelo y sedal).

1.3.3 (Una actividad divertida) Escriba los números 1, 2, 3 y 4 en orden en una tarjeta. Lleve la tarjeta a un lugar concurrido (por ejemplo, un restaurante, una biblioteca, una universidad) y pida a un mínimo de 30 personas que miren la tarjeta y seleccionen uno de los dígitos aleatoriamente. Apunte sus respuestas.

- Si las personas pueden pensar «aleatoriamente», ¿qué fracción aproximada de las personas responderían con el dígito 1? ¿Y el 2? ¿Y el 3? ¿Y el 4?
- ¿Qué fracción de las personas preguntadas respondió con el dígito 1? ¿Y el 2? ¿Y el 3? ¿Y el 4?
- ¿Sugieren algo los resultados sobre la capacidad de la gente de escoger aleatoriamente?

1.3.4 Considere una población formada por 600 individuos con identificadores únicos: 001, 002, ..., 600. Utilice la siguiente cadena de dígitos aleatorios para seleccionar una muestra aleatoria simple de 5 individuos. Escriba los indicadores de los individuos seleccionados en su muestra.

728121876442121593787803547216596851

1.3.5 (Ejercicio de muestreo) Observe el conjunto de 100 elipses en la figura que acompaña, que se puede imaginar como una representación de una población natural de un organismo mítico *C. ellipticus*. Las elipses han recibido los números identificativos 00, 01, ..., 99 por conveniencia en el muestreo. Ciertos individuos de *C. ellipticus* son mutantes y presentan dos cerdas en la cola.

- Utilice su *juicio* para escoger una muestra de tamaño 10 de la población que piense que sea representativa de la población completa. Note el número de mutantes en la muestra.

(b) Utilice *dígitos aleatorios* (de la Tabla 1 o de su calculadora o computador) para escoger una muestra aleatoria de tamaño 10 de la población y vea el número de mutantes en la muestra.

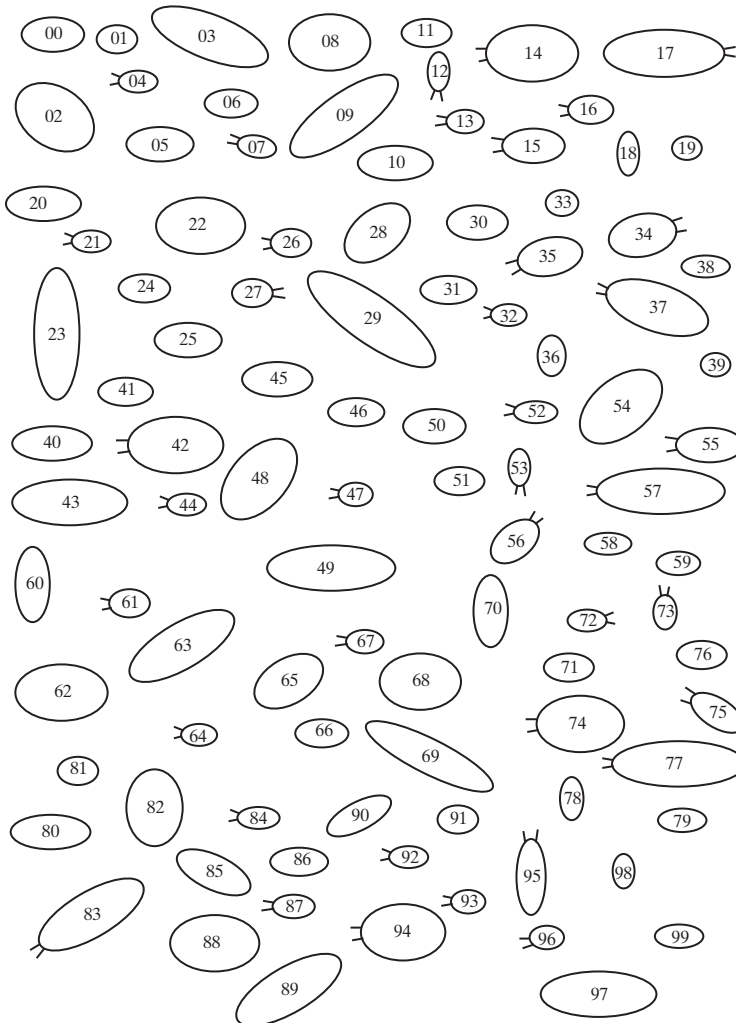
1.3.6 (Ejercicio de muestreo) Observe de nuevo el conjunto de 100 elipses.

(a) Utilice dígitos aleatorios (de la Tabla 1, calculadora o computador) para escoger una muestra aleatoria de tamaño 5 de la población y vea el número de mutantes en la muestra.

(b) Repita el apartado (a) nueve veces más, para un total de 10 muestras (algunas de las 10 muestras se pueden solapar).

Para facilitar la recopilación de los resultados de la clase completa, presente sus resultados en el siguiente formato:

Número de mutantes	No mutantes	Frecuencia (número de muestras)
0	5	
1	4	
2	3	
3	2	
4	1	
5	0	
		Total: 10



Notas

1. Nicolle, J. (1961). *Louis Pasteur: The Story of His Major Discoveries*. New York: Basic Books, p. 170. © 1961 de Jacques Nicolle. © 1961 Traducción al inglés Hutchinson & Co. (Publishers) Ltd. Reimpreso con permiso de Perseus Books Group.
2. Mizutani, T. y Mitsuoka, T. (1979). Effect of intestinal bacteria on incidence of liver tumors in gnotobiotic C3H/He male mice. *Journal of the National Cancer Institute* **63**, 1365-1370.
3. Tripepi, R. R. y Mitchell, C. A. (1984). Metabolic response of river birch and European Birch roots to hypoxia. *Plant Physiology* **76**, 31-35. Datos iniciales cortesía de los autores.
4. Adaptado de Potkin, S. G., Cannon, H. E., Murphy, D. L. y Wyatt, R. J. (1978). Are paranoid schizophrenics biologically different from other schizophrenics? *New England Journal of Medicine* **298**, 61-66. Los datos son aproximados, se han reconstruido de los histogramas e información resumen dada por Potkin *et al.* Reimpreso con permiso del *New England Journal of Medicine*.
5. Wolfson, J. L. (1987). Impact of *Rhizobium* nodules on *Sitona hispidulus*, the clover root curculio. *Entomologia Experimentalis et Applicata* **43**, 237-243. Datos cortesía del autor. El experimento real incluía 12 placas.
6. Webb, P. (1981). Energy expenditure and fat-free mass in men and women. *American Journal of Clinical Nutrition* **34**, 1816-1826.
7. El titular apareció en la página 2 de la edición dominical del *The New York Times*, 16 de julio de 1911.
8. Allen, L. S. y Gorski, R. A. (1992). Sexual orientation and the size of the anterior commissure in the human brain. *Proceedings of the National Academy of Science* **89**, 7199-7202. Los datos son aproximados y se han reconstruido de los diagramas de puntos e información resumen dada por Allan y Gorski. Respecto al primer asunto mencionado en el Ejemplo 1.2.2, los autores eran conscientes del efecto que las dos observaciones más grandes podían tener en sus conclusiones y calcularon el promedio para los hombres homosexuales una segunda vez, tras borrar esos dos valores. Para el segundo asunto, los autores calcularon los promedios de los que tenían SIDA y los que no en cada grupo de hombres. Obtuvieron que el SIDA está asociado con áreas AC menores, no mayores, de forma que cuando solo se comparan personas sin SIDA, la diferencia entre hombres homosexuales y heterosexuales es incluso mayor que la diferencia obtenida en el conjunto completo de datos.
9. Hakansson, K., Rovio, S., Helkala, E. L., Vilska, A. R., Winblad, B., Soininen, H., Nissinen, A., Mohammed, A. H. y Kivipelto, M. (2009). Association between mid-life marital status and cognitive function in later life: Population based cohort study. *BMJ* **339**, b2462.
10. Bradstreet, T. E. (1992). Favorite data sets from early phases of drug research-part 2. *Proceedings of the Section on Statistical Education of the American Statistical Association*, 219-223.
11. Benson, H. y Friedman, R. (1996). Harnessing the power of the placebo effect and renaming it «remembered wellness». En *Annual Review of Medicine* **47**, 193-199. Annual Reviews, Inc., Palo Alto, Calif.
12. Sandler, A. D., Sutton, K. A., DeWeese, J., Girardi, M. A., Sheppard, V. y Bodfish, J. W. (1999). Lack of benefit of a single dose of synthetic human secretin in the treatment of autism and pervasive developmental disorder. *New England Journal of Medicine* **341**, 1801-1806. La mejora en el grupo del placebo fue algo mejor que la mejora en el grupo del secretin para la variable de respuesta del cambio en la puntuación total, pero el *P* valor de la diferencia fue de 0,11.
13. Butler, C. y Steptoe, A. (1986). Placebo response: An experimental study of asthmatic volunteers, *British Journal of Clinical Psychology* **25**, 173-183.
14. Barsamian, E. M. (1977). The rise and fall on internal mammary artery ligation in the treatment of angina pectoris and the lessons learned: En Bunker, J. P., Barnes, B. A. y Mosteller, F. (eds.), *Costs, Risks, and Benefits of Surgery*. New York: Oxford University Press, pp. 212-220.
15. Chalmers, T. C., Celano, P., Sacks, H. S. y Smith, H. (1983). Bias in treatment assignment in controlled clinical trials. *New England Journal of Medicine* **309**, 1358-1361.
16. The Coronary Drug Project Research Group (1980). Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *New England Journal of Medicine* **303**, 1038-1041. Se midieron varias variables de cada sujeto al comienzo del experimento. Ajustando los efectos de las covariaciones en el grupo del placebo solo se reduce ligeramente la diferencia en las tasas de mortalidad entre los que se adhirieron y los que no se adhirieron. Por tanto, las diferencias en la salud global solo explican una pequeña parte de las diferencias en las tasas de mortalidad de los que se adhirieron y los que no se adhirieron.
17. Diehl, H. S., Baker, A. B. y Cowan, D. W. (1938). Cold vaccines: An evaluation based on a controlled study. *Journal of the American Medical Association* **111**, 1168-1173.
18. Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. y Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer* **34**, 585-612.
19. Sacks, H., Chalmers, T. C. y Smith, H. (1982). Randomized versus historical controls for clinical trials. *American Journal of Medicine* **72**, 233-240.

20. Dublin, L. I. (1957). *Water fluoridation: Facts, not myths*. New York: Public Affairs Committee, Inc.
21. Sandler, R. S., Zorich, N. L., Filloon, T. G., Wiseman, H. B., Lietz, D. J., Brock, M. H., Royer, M. G. y Miday, R. K. (1999). Gastrointestinal symptoms in 3181 volunteers ingesting snack foods containing olestra or triglycerides. A 6-week randomized, placebo-controlled trial. *Annals of Internal Medicine* **130**, 253-261.
22. Moertel, C. G., Fleming, T. R., Creagan, E. T., Rubin, J., O'Connell, M. J. y Ames, M. M. (1985). High-dose vitamin C versus placebo in the treatment of patients with advanced cancer who have had no prior chemotherapy. *New England Journal of Medicine* **312**, 137-141.
23. Pauling, L. y Cameron, E. (1976). Supplemental ascorbate in the supportive treatment of cancer: Prolongation of survival times in terminal human cancer. *Proceedings of the National Academy of Sciences* **73**, 3685-3789.
24. *Cleveland Plain Dealer*, 23 de octubre de 1997, página 15-A.
25. Huber, Ann M. (2005). Moisture requirements for the germination of early seedling survival of *Cirsium loncholepis*. Master's thesis in Biological Sciences, California Polytechnic State University.
26. Comunicación personal de L. Vredevoe con respecto a un proyecto de investigación en marcha (2009).
27. Parks, N. J., Krohn, K. A., Mathis, C. A., Chasko, J. H., Geiger, K. R., Gregor, M. E. y Peek, N. F. (1981). Nitrogen-13-labelled nitrite and nitrate: Distribution and metabolism after intratracheal administration. *Science* **212**, 58-61.
28. *Cleveland Plain Dealer*, 25 de junio de 1991, página 3-A.
29. Hull, H. F., Bettinger, C. J., Gallaher, M. M., Keller, N. M., Wilson, J. y Mertz, G. J. (1988). Comparison of HIV-antibody prevalence in patients consenting to and declining HIV-antibody testing in an STD clinic. *Journal of the American Medical Association*, **260**, 935-938.
30. Hoover, D. W. y Milich, R. (1994). Effects of sugar ingestion expectancies on mother-child interactions. *Journal of Abnormal Psychiatry* **22**, 501-515.
31. Krummel, D. A., Seligson, F. H. y Guthrie, H. A. (1996). Hyperactivity: Is candy causal? *Critical Reviews in Food Science and Nutrition* **36**, 31-47.

DESCRIPCIÓN DE MUESTRAS Y POBLACIONES

2

Objetivos

En este capítulo estudiaremos cómo describir datos. En particular,

- mostraremos cómo se utilizan las distribuciones de frecuencias para realizar diagramas de barras e histogramas;
- compararemos la media y la mediana como medidas de tendencia central;
- mostraremos cómo se construyen e interpretan diversos gráficos como diagramas de puntos, diagramas de cajas y diagramas de dispersión;
- compararemos varias medidas de variabilidad con especial énfasis en la desviación típica;
- examinaremos cómo afectan a las distribuciones las transformaciones de variables;
- consideraremos la relación entre poblaciones y muestras.

2.1 Introducción

La estadística es la ciencia de analizar y extraer información de los datos. En esta sección presentaremos alguna terminología y notación para trabajar con datos.

Variables

Empezaremos con el concepto de **variable**. Una variable es una característica de una persona o cosa a la que se le puede asignar un número o una categoría. Por ejemplo, el grupo sanguíneo (A, B, AB, O) y la edad son dos variables que podrían ser medidas en una persona.

El grupo sanguíneo es un ejemplo de **variable categórica**. Una variable categórica es una variable que almacena a cuál de una serie de categorías pertenece una persona o cosa. Ejemplos de variables categóricas pueden ser:

- El grupo sanguíneo de una persona: A, B, AB, O.
- El sexo de un pez: macho o hembra.
- El color de una flor: rojo, rosa, blanco.
- La forma de una semilla: arrugada, lisa.

En algunas variables categóricas, las categorías se pueden disponer en un orden que tiene sentido. Esas variables se denominan **ordinales**. Por ejemplo, la respuesta de un paciente a una terapia puede ser ninguna, parcial o completa.

La edad es un ejemplo de **variable numérica**. Una variable numérica es una variable que almacena la cantidad de algo. Una **variable continua** es una variable numérica que se mide en una escala continua. Ejemplos de variables continuas pueden ser:

- El peso de un bebé.
- La concentración de colesterol en un espécimen de sangre.
- La densidad óptica de una solución.

Una variable como el peso es continua porque, en principio, dos pesos pueden estar arbitrariamente próximos. Algunos tipos de variables numéricas no son continuas, pero se pueden situar en una escala discreta, con espacios entre los posibles valores. Una **variable discreta** es una variable numérica para la que se puede dar una lista de sus posibles valores. Por ejemplo, el número de huevos en el nido de un pájaro es una variable discreta porque solo los valores 0, 1, 2, 3, ..., son posibles. Otros ejemplos de variables discretas pueden ser:

- El número de colonias de bacterias en una placa de Petri.
- El número de ganglios linfáticos cancerosos detectados en un paciente.
- La longitud en parejas de bases de un segmento de ADN.

La distinción entre variables continuas y discretas no es rígida. Después de todo, todas las medidas físicas se redondean. El peso de un novillo se mide en kilogramos, el de una rata en gramos y el de un insecto en miligramos. Estrictamente hablando, la escala de una medida es siempre discreta. La escala continua puede verse como una aproximación a la escala real de las medidas.

Unidades observacionales

Cuando se tiene una muestra de n personas o cosas y se miden una o más variables sobre ellas, dichas personas o cosas se denominan **unidades observacionales** o casos. A continuación presentamos algunos ejemplos de muestras.

Muestra	Variable	Unidad observacional
150 bebés nacidos en un cierto hospital	Peso del bebé (kg)	Un bebé
73 palomillas <i>Cecropia</i> atrapadas en una trampa	Sexo	Una palomilla
81 plantas que forman la progenie de un único cruce parental	Color de las flores	Una planta
Colonias de bacterias en cada una de seis placas de Petri	Número de colonias	Una placa de Petri

Notación para variables y observaciones

Adoptaremos una convención rotacional para distinguir entre una variable y un valor observado de esa variable. Designaremos a las variables con letras mayúsculas, como la Y , y las observaciones (es decir, los datos) con letras minúsculas, como la y . Así pues, distinguimos por ejemplo, entre Y = peso del bebé (la variable) e $y = 3$ kg (la observación). Esta distinción será útil para la explicación de algunas ideas fundamentales relativas a la variabilidad.

Ejercicios 2.1.1-2.1.4

En cada una de las situaciones de los ejercicios 2.1.1-2.1.4: (i) Identifique la variable o variables del estudio; (ii) para cada variable, indique el tipo de variable (es decir, categórica y ordinal, discreta, etc.); (iii) identifique la unidad observacional (es decir, lo que se muestrea), y (iv) identifique el tamaño de la muestra.

2.1.1

- Un paleontólogo midió la anchura (en mm) del último molar en 36 especímenes del mamífero extinto *Acrophitecus rigidus*.
- Se apuntaron el peso al nacer, fecha de nacimiento y raza de la madre de 65 bebés.

2.1.2

- Un médico midió la altura y el peso de 37 niños.
- Durante una donación de sangre, un banco de sangre ofreció comprobar el colesterol de cualquiera que do-

nara sangre. Un total de 129 personas donaron sangre. Se apuntaron, para cada una de ellas, el grupo sanguíneo y el nivel de colesterol.

2.1.3

- Un biólogo midió el número de hojas en 25 plantas.
- Un médico anotó el número de ataques de un grupo de personas con epilepsia severa en un periodo de ocho semanas.

2.1.4

- Un conservacionista apuntó el tiempo (despejado, parcialmente nublado, nublado, lluvioso) y el número de coches aparcados a mediodía en un sendero durante un periodo de 18 días.
- Un enólogo midió el pH y el contenido residual de azúcar (g/l) de siete barriles de vino.

2.2 Distribuciones de frecuencias

Un primer paso para entender el significado de un conjunto de datos de una variable dada consiste en explorar dichos datos y describirlos en forma resumida. En este capítulo presentaremos tres aspectos mutuamente complementarios de la descripción resumida de datos: las distribuciones de frecuencias, las medidas de tendencia central y las medidas de dispersión. Estas herramientas nos informarán de la forma, el centro y la dispersión de los datos.

Una **distribución de frecuencias** es simplemente una presentación de la **frecuencia**, o número de apariciones de cada valor en el conjunto de datos. La información se puede presentar en forma de tabla o, de forma más vistosa, mediante un gráfico. Un **diagrama de barras** es un gráfico simple que muestra las categorías que puede tomar una variable categórica y el número de observaciones de cada categoría en los datos de la muestra. Presentamos a continuación dos ejemplos de distribuciones de frecuencia de datos categóricos.

Ejemplo 2.2.1 Color de flores de pascua

Las flores de pascua pueden ser rojas, rosas o blancas. En una investigación sobre el mecanismo hereditario de control del color, se clasificó por color una progenie de 182 de cierto cruce parental¹. El diagrama de barras de la Figura 2.2.1 muestra de forma visual los resultados de la Tabla 2.2.1. ■

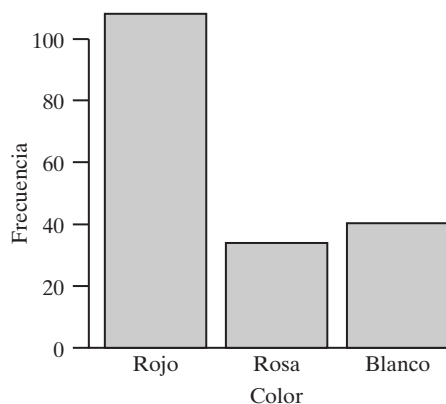


Figura 2.2.1 Diagrama de barras del color de 182 flores de pascua

Tabla 2.2.1 Color de 182 flores de pascua	
Color	Frecuencia (número de plantas)
Rojo	108
Rosa	34
Blanco	40
Total	182

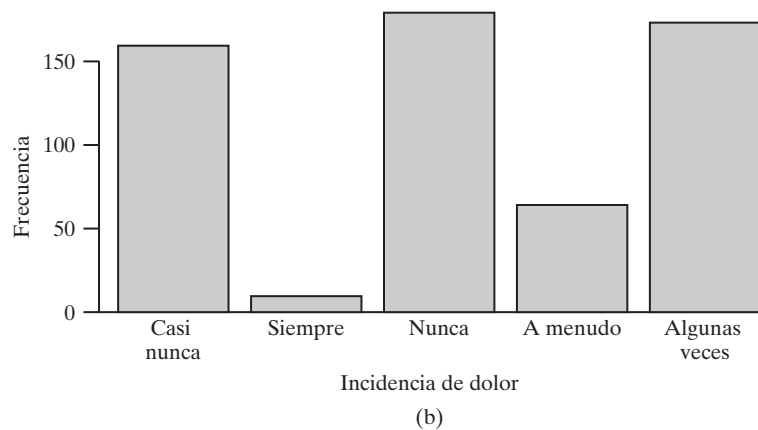
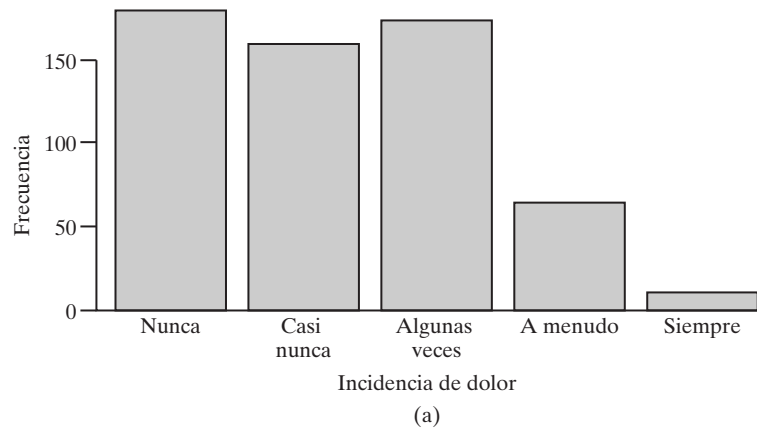
Ejemplo 2.2.2 Mochilas escolares y dolores de cuello

Unos fisiólogos australianos consideraron la posibilidad de que llevar una mochila cargada con libros pesados era una causa de dolores de cuello en adolescentes, por lo que preguntaron a una muestra de 585 chicas adolescentes con qué frecuencia tenían dolores de cuello al llevar su mochila (nunca, casi nunca, algunas veces, a menudo, siempre). La Tabla 2.2.2 muestra un resumen de los resultados obtenidos, que se muestran en forma de diagrama de barras en la Figura 2.2.2(a)². Como la variable incidencia es una variable categórica ordinal, nuestras tablas y gráficas deben respetar su ordenación natural. La Figura 2.2.2(b) muestra los mismos datos pero con las categorías en orden alfabético (una opción por defecto en muchos *softwares*), lo que oscurece la información de los datos. ■

Un **diagrama de puntos** es un gráfico sencillo que se puede utilizar para mostrar la distribución de una variable numérica cuando el tamaño de la muestra es pequeño. Para realizar un diagrama de puntos, se dibuja una recta numérica que abarque el intervalo de los datos y después se pone un punto sobre la línea numérica en el lugar correspondiente a cada observación, como se muestra en el siguiente ejemplo.

Tabla 2.2.2 Dolor de cuello asociado a llevar una mochila escolar

Incidencia	Frecuencia (número de chicas)
Nunca	179
Casi nunca	159
Algunas veces	173
A menudo	64
Siempre	10
Total	585

**Figura 2.2.2** (a) Diagrama de barras de la incidencia de dolor de cuello indicada por 585 adolescentes. (b) Los mismos datos, pero con las categorías en orden aleatorio**Ejemplo 2.2.3** Mortalidad infantil

La Tabla 2.2.3 muestra la tasa de mortalidad infantil (muertes infantiles por cada 100 nacimientos vivos) en 12 países de Sudamérica en 2009³. La distribución se muestra en la Figura 2.2.3. ■

Cuando dos o más observaciones tienen el mismo valor, se sitúan los puntos del diagrama de barras uno encima de otro. Esto produce un efecto similar al de las barras en un diagrama de barras. Si creamos barras en vez de pilas de puntos, entonces tenemos un **histograma**. Un histograma es como un diagrama de barras, excepto porque un histograma muestra una variable numérica, lo que significa que hay un orden natural y una escala para la variable. En un diagrama de barras el espacio entre las barras (si existe) es arbitrario, ya que los datos que se muestran son categóricos. En un histograma, la escala de la variable determina la situación de las barras. El ejemplo que sigue muestra un diagrama de puntos y un histograma de una distribución de frecuencias.

Ejemplo 2.2.4 Tamaño de camada de cerdas

Un grupo de 32 cerdas de 2 años de edad de la misma raza (3/4 Duroc, 1/4 Yorkshire) se aparearon con cerdos Yorkshire. Se apuntó el número de cochinitos supervivientes de cada cerda tras 21 días⁴. Los resultados se muestran en la Tabla 2.2.4 y en forma gráfica de diagrama de puntos en la Figura 2.2.4 y en un histograma en la Figura 2.2.5. ■

Tabla 2.2.3 Mortalidad infantil en 12 países de Sudamérica

País	Tasa de mortalidad infantil
Argentina	11,4
Bolivia	44,7
Brasil	22,6
Chile	7,7
Colombia	18,9
Ecuador	20,9
Guayana	30,0
Paraguay	24,7
Perú	28,6
Surinam	18,8
Uruguay	11,3
Venezuela	26,5



Figura 2.2.3 Diagrama de puntos de la mortalidad infantil en 12 países de Sudamérica

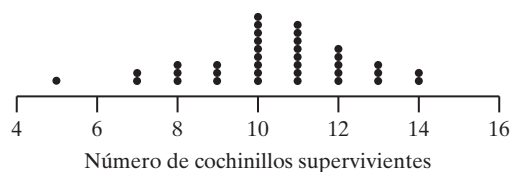


Figura 2.2.4 Diagrama de puntos del número de cochinillos supervivientes de 36 cerdas

Tabla 2.2.4 Número de cochinillos supervivientes de 36 cerdas

Número de cochinillos	Frecuencia (número de cerdas)
5	1
6	0
7	2
8	3
9	3
10	9
11	8
12	5
13	3
14	2
Total	36

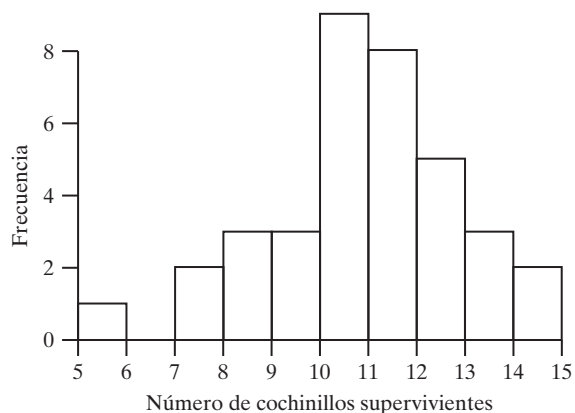


Figura 2.2.5 Histograma del número de cochinillos supervivientes de 36 cerdas

Frecuencia relativa

La escala de frecuencia se sustituye a menudo por una escala de **frecuencia relativa**:

$$\text{Frecuencia relativa} = \frac{\text{Frecuencia}}{n}$$

La escala de frecuencia relativa es útil si se van a presentar juntos varios conjuntos de datos de diferentes tamaños (n 's) para su comparación. Como otra opción, una frecuencia relativa se puede expresar como una frecuencia en porcentaje. La forma de la gráfica no se altera por la selección de la escala de frecuencias, como se muestra en el siguiente ejemplo.

Ejemplo 2.2.5 Color de flores de pascua

La distribución del color de las flores de pascua del Ejemplo 2.2.1 se expresa como frecuencia, frecuencia relativa y frecuencia en porcentaje en la Tabla 2.2.5 y la Figura 2.2.6. ■

Color	Frecuencia	Frecuencia relativa	Frecuencia en porcentaje
Rojo	108	0,59	59
Rosa	34	0,19	19
Blanco	40	0,22	22
Total	182	1,00	100

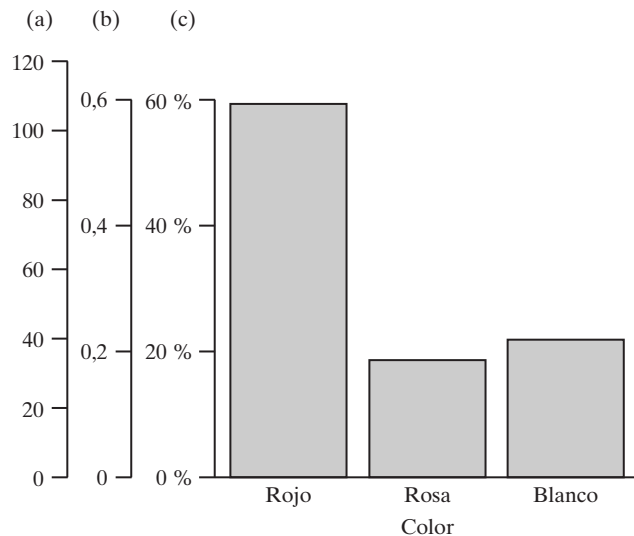


Figura 2.2.6 Diagrama de barras de los colores de flores de pascua en tres escalas: (a) frecuencia; (b) frecuencia relativa; (c) frecuencia en porcentaje

Distribuciones de frecuencias agrupadas

En los ejemplos anteriores, las distribuciones de frecuencia simples sin agrupar han proporcionado resúmenes concisos de los datos. En muchos conjuntos de datos, es necesario agrupar dichos datos para condensar la información adecuadamente. (Generalmente, esto sucede en el caso de variables continuas). El ejemplo que sigue muestra una distribución de frecuencias agrupadas.

Ejemplo 2.2.6 CFK en el suero

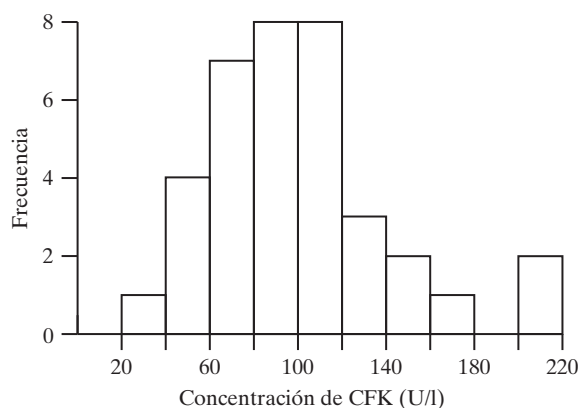
La creatina fosfoquinasa es una enzima relacionada con las funciones de los músculos y del cerebro. Como parte de un estudio para determinar la variación natural de concentración de CFK, se extrajo sangre de 36 voluntarios varones. La Tabla 2.2.6 muestra sus concentraciones en suero de CFK (medidas en U/l)⁵. La Tabla 2.2.7 muestra dichos datos agrupados en **clases**. Por ejemplo, la frecuencia de la clase [20, 40) (todos los valores en el intervalo $20 \leq y < 40$) es 1, lo que significa que solo un valor de CFK estaba en ese intervalo. La distribución de datos agrupados se muestra como histograma en la Figura 2.2.7. ■

Tabla 2.2.6 Valores de CFK en el suero de 36 hombres

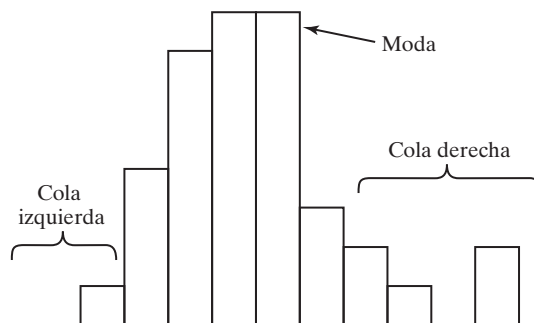
121	82	100	151	68	58
95	145	64	201	101	163
84	57	139	60	78	94
119	104	110	113	118	203
62	83	67	93	92	110
25	123	70	48	95	42

Tabla 2.2.7 Distribución de frecuencias de los valores de CFK en el suero de 36 hombres

CFK en el suero (U/l)	Frecuencia (número de hombres)
[20, 40)	1
[40, 60)	4
[60, 80)	7
[80, 100)	8
[100, 120)	8
[120, 140)	3
[140, 160)	2
[160, 180)	1
[180, 200)	0
[200, 220)	2
Total	36

**Figura 2.2.7** Histograma de la concentración de CFK en suero de 36 hombres

Una distribución de frecuencias agrupadas debe mostrar los aspectos esenciales de los datos. Por ejemplo, el histograma de la Figura 2.2.7 muestra que el valor promedio de CFK está alrededor de 100 U/l, y que la mayoría de los valores están entre 60 y 140 U/l. Además, el histograma muestra la *forma* de la distribución. Nótese que los valores de CFK se acumulan alrededor de un pico central, o **moda**. A cada lado de la moda, las frecuencias disminuyen hasta formar las **colas** de la distribución. Estas características de forma se indican en la Figura 2.2.8. La distribución de CFK no es simétrica, sino que está un poco **sesgada hacia la derecha**, lo que quiere decir que la cola derecha está más estirada que la izquierda*.

**Figura 2.2.8** Características de forma de la distribución de CFK

* Para ayudar a recordar qué cola de una distribución sesgada es la mayor, véase el sesgo como un estiramiento. ¿Qué lado de la distribución está más estirado hacia el exterior del centro? Una distribución sesgada a la derecha es aquella en la que la cola derecha está más estirada que la izquierda.

Al realizar un histograma, es necesario decidir cuántas clases hay que tener y la anchura de dichas clases. Si utilizamos *software* estadístico para generar un histograma, el programa escogerá por nosotros el número de clases y la anchura de dichas clases, pero la mayor parte del *software* permite al usuario cambiar el número de clases y especificar su anchura. Si el conjunto de datos es grande y muy disperso, es buena idea ver más de un histograma de los datos, tal como se hace en el Ejemplo 2.2.7.

Ejemplo 2.2.7 Alturas de estudiantes

Se preguntó la altura a una muestra de 510 estudiantes universitarios. Nótese que no fueron medidos, sino que solo dijeron lo que medían⁶. La Figura 2.2.9 muestra la distribución de los valores indicados por los estudiantes, utilizando 7 clases y una anchura de clase de 3 (pulgadas). Utilizando solo 7 clases, la distribución parece ser razonablemente simétrica, con un único pico alrededor de 66 pulgadas.

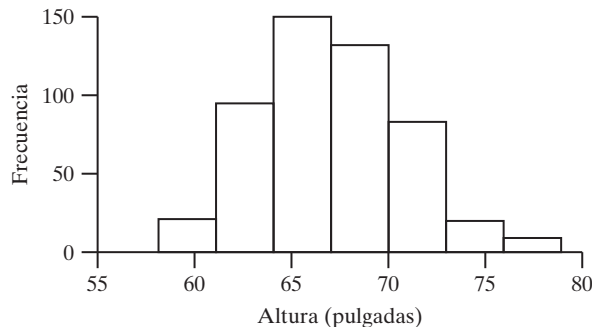


Figura 2.2.9 Alturas de estudiantes, utilizando 7 clases (anchura de clase = 3)

La Figura 2.2.10 muestra los datos de altura pero en un histograma que utiliza 18 clases y una anchura de clase de 1,1. Esta presentación de los datos muestra dos modas, una para los hombres y otra para las mujeres.

La Figura 2.2.11 muestra de nuevo los datos de altura, esta vez utilizando 37 clases, cada una de ellas de anchura 0,5. Al utilizar un número tan grande de clases, la distribución toma un aspecto irregular. En este caso, puede observarse un patrón alternante entre clases con muchas observaciones y clases con pocas observaciones. En la mitad de la distribución podemos ver que muchos estudiantes indicaron una altura de 63 pulgadas, y pocos indicaron 63,5 pulgadas, muchos que indicaron 64 pulgadas, y así sucesivamente. ¡Parece que la mayoría de los estudiantes redondearon a la pulgada más cercana!

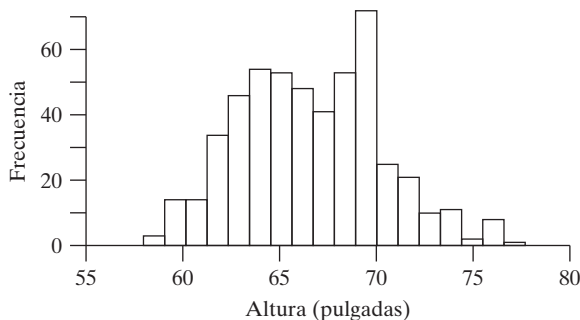


Figura 2.2.10 Alturas de estudiantes, utilizando 18 clases (anchura de clase = 1,1)

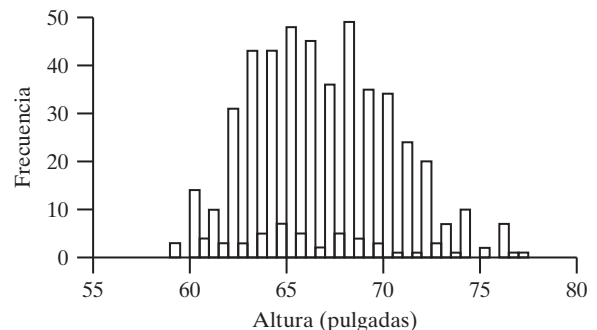


Figura 2.2.11 Alturas de estudiantes, utilizando 37 clases (anchura de clase = 0,5)

Interpretación de las áreas en un histograma

Un histograma se puede ver de dos formas. La parte superior de las barras esboza la forma de la distribución. Pero las *áreas* dentro de las barras tienen también un significado. El área de cada barra es proporcional a la correspondiente frecuencia. Por tanto, el área de una o varias barras se puede interpretar que expresa el número de observaciones en las clases representadas por las barras. Por ejemplo, la Figura 2.2.12 muestra un histograma de la distribución de CFK del Ejemplo 2.2.6. El área sombreada es el 42 % del área total de todas las barras. De acuerdo con esto, el 42 % de los valores de CFK están en las clases correspondientes. Es decir, 15 de 36 o el 42 % de los valores están entre 60 U/l y 100 U/l*.

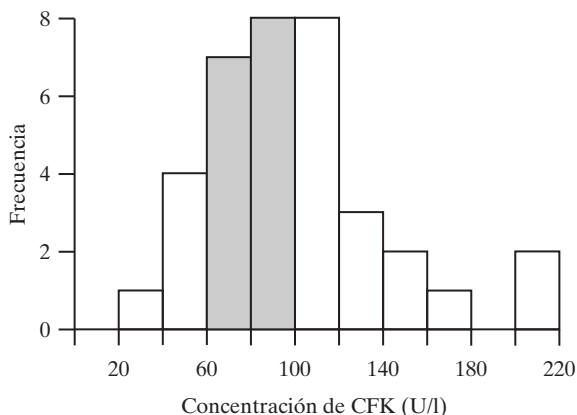


Figura 2.2.12 Histograma de la distribución de CFK. El área sombreada es el 42 % del área total y representa el 42 % de las observaciones

La interpretación de las áreas en el histograma es una idea simple, pero importante. En nuestro trabajo posterior con distribuciones encontraremos que dicha idea es indispensable.

Formas de distribuciones

Al estudiar un conjunto de datos, hay que describir la forma, el centro y la dispersión de la distribución. En esta sección nos concentramos en las formas de las distribuciones de frecuencias e ilustramos parte de la diversidad de distribuciones que se encuentran en las Ciencias de la Vida. La forma de una distribución se puede indicar mediante una curva suave que se aproxime al histograma, como se muestra en la Figura 2.2.13.

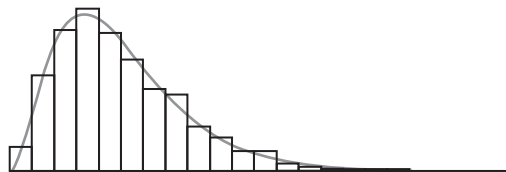


Figura 2.2.13 Aproximación al histograma mediante una curva suave

La Figura 2.2.14 muestra algunas formas de distribuciones. Una forma común para datos biológicos es la **unimodal** (que tiene una moda), y está algo sesgada hacia la derecha, como en (c). Aparecen también distribuciones con forma aproximada de campana, como en (a). Algunas veces una distribución es simétrica, pero se diferencia de una campana porque tiene colas largas, como en (b). Las formas sesgadas hacia la izquierda (b) y exponencial (e) son menos comunes. La **bimodalidad**, como en (f), puede indicar la existencia de dos subgrupos distintos de unidades observacionales.

* Estrictamente hablando, entre 60 U/l y 99 U/l, ambas inclusive.

Nótese que las características de forma que estamos enfatizando, como el número de modas y el grado de simetría, son *insensibles a la escala*, es decir, no son afectadas por la selección arbitraria de las escalas horizontal y vertical al dibujar la distribución. Por el contrario, una característica como la de si la distribución aparece corta y gruesa o alta y fina sí queda afectada por la forma en la que se dibuja la distribución y, por tanto, no es una característica inherente de una variable biológica.

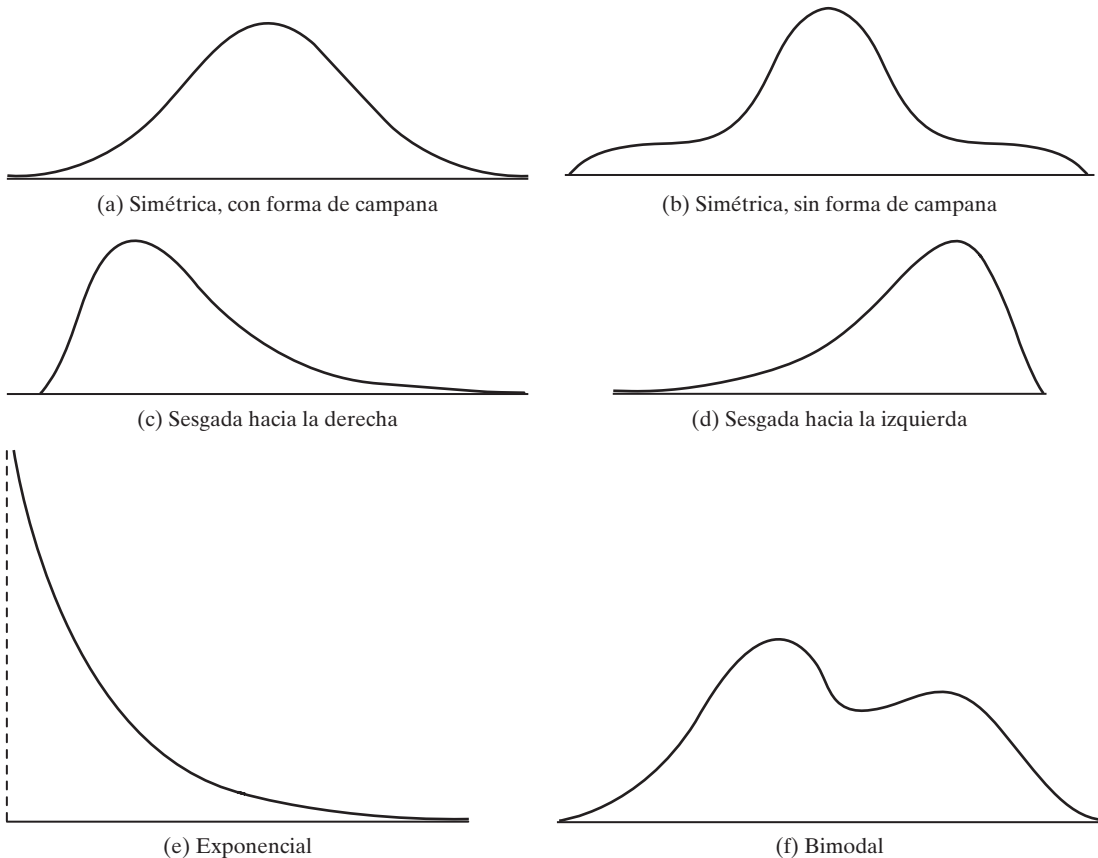


Figura 2.2.14 Formas de distribuciones

Los tres ejemplos que siguen ilustran distribuciones de frecuencias en biología con diversas formas. En el primer ejemplo, la forma de la distribución proporciona evidencia de que la estructura analizada es en verdad biológica en vez de no biológica.

Ejemplo 2.2.8 Microfósiles

En 1977, los paleontólogos descubrieron estructuras fósiles microscópicas, semejantes a algas, en rocas de 3.500 millones de años de antigüedad. La cuestión primordial era si esas estructuras tenían origen biológico. Una línea de argumentación se centró en la distribución de su tamaño, que se muestra en la Figura 2.2.15. Esta distribución, que es unimodal y bastante simétrica, recuerda a las de poblaciones de microbios conocidas, pero no a la de estructuras no biológicas conocidas⁷.

Ejemplo 2.2.9 Tiempo entre descargas de células

Un neurobiólogo observó descargas en células de músculo de rata que se cultivaron junto con células nerviosas. La Figura 2.2.16 muestra la distribución de los intervalos de tiempo entre 308 descargas sucesivas. Nótese la forma exponencial de la distribución⁸.

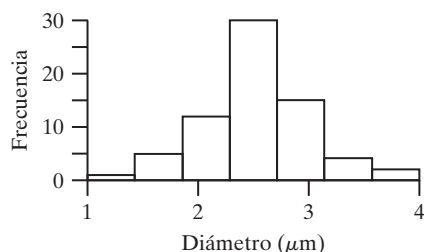


Figura 2.2.15 Tamaños de microfósiles

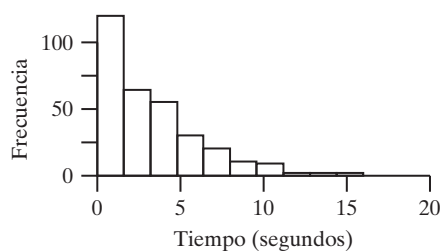


Figura 2.2.16 Intervalos de tiempo entre descargas eléctricas en células de músculo de rata

Ejemplo 2.2.10 Peso del cerebro

En 1888, P. Topinard publicó datos sobre los pesos de los cerebros de cientos de hombres y mujeres franceses. Las Figuras 2.2.17(a) y (b) muestran los datos para los hombres y las mujeres. La distribución para los hombres es bastante simétrica y con forma de campana. La distribución para las mujeres es algo sesgada hacia la derecha. La parte (c) de la figura muestra la distribución del peso del cerebro conjunta para hombres y mujeres. Esta distribución es ligeramente bimodal⁹. ■

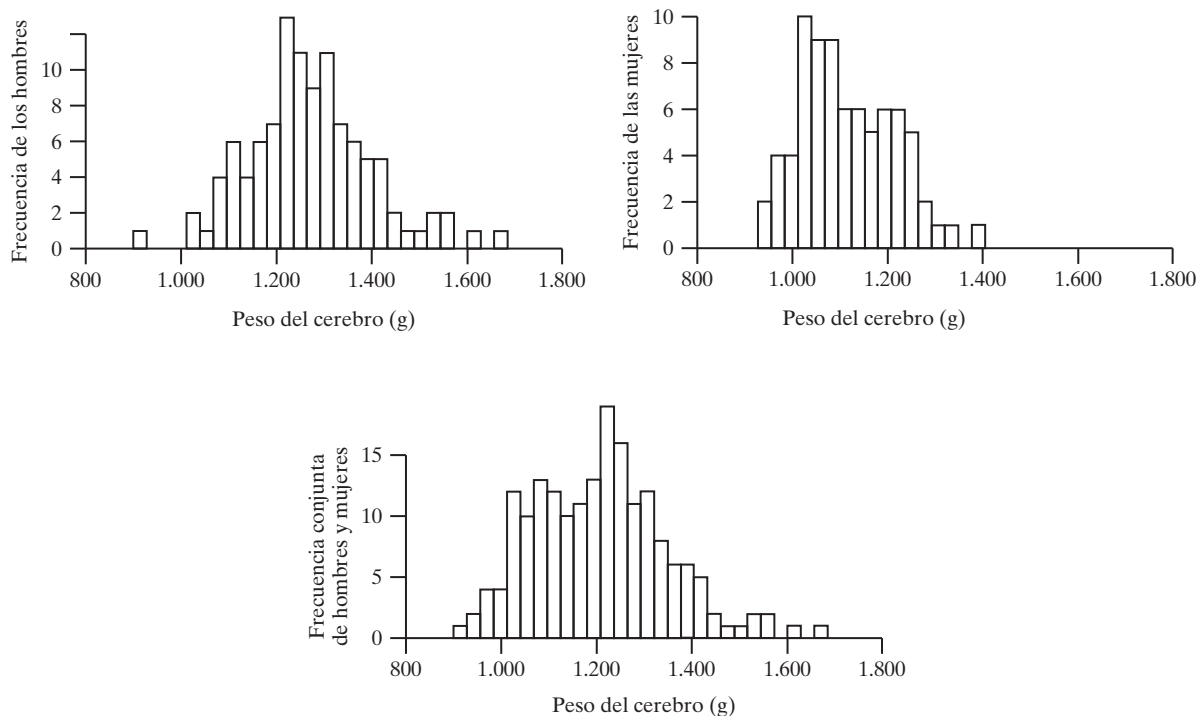


Figura 2.2.17 Peso del cerebro

Fuentes de variación

Al interpretar datos biológicos, es útil tener en cuenta las fuentes de variabilidad. La variación entre observaciones en un conjunto de datos refleja a veces los efectos combinados de varios factores subyacentes. Los dos ejemplos siguientes ilustran estas situaciones.

Ejemplo 2.2.11 Pesos de semillas

En un experimento clásico para distinguir la influencia del entorno en la genética, un genetista pesó semillas de la judía alada *Phaseolus vulgaris*. La Figura 2.2.18 muestra las distribuciones de pesos de (a) 5.494 semillas de un lote de semillas comercial y (b) 712 semillas de una línea altamente endogámica proveniente de una sola semilla del lote original. La variabilidad en (a) es debida a factores tanto genéticos como del entorno. En (b), como las plantas son genéticamente casi idénticas, la variabilidad de los pesos es debida fundamentalmente a la influencia del entorno¹⁰. Por tanto, hay menos variabilidad en la línea endogámica. ■

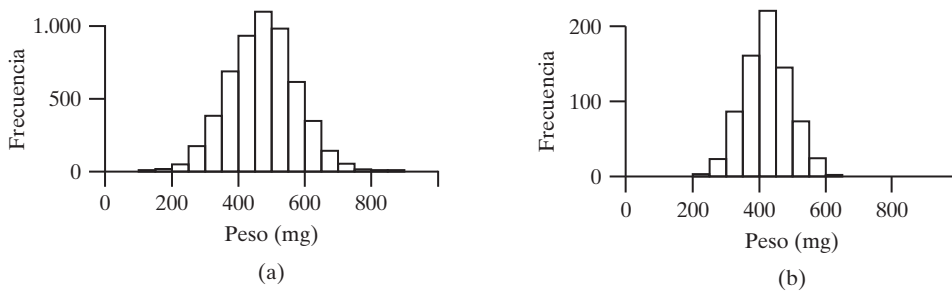


Figura 2.2.18 Pesos de semillas de judía alada: (a) de una población abierta; (b) de una población endogámica

Ejemplo 2.2.12 ALT en el suero

La alanina aminotransferasa (ALT) es una enzima que se encuentra en la mayoría de los tejidos humanos. La parte (a) de la Figura 2.2.19 muestra las concentraciones en suero de ALT en 129 voluntarios adultos. Posibles fuentes de variabilidad entre las medidas pueden ser:

1. Interindividuos:
 - (a) Genéticas.
 - (b) Del entorno.
2. Intraindividuos:
 - (a) Cambios biológicos en el tiempo.
 - (b) Imprecisión analítica en el ensayo.

El efecto de la última fuente (variación analítica) se puede ver en la parte (b) de la Figura 2.2.19, que muestra la distribución de frecuencias de 109 ensayos del *mismo* espécimen de suero. La figura indica que el ensayo de ALT es bastante impreciso¹¹. ■

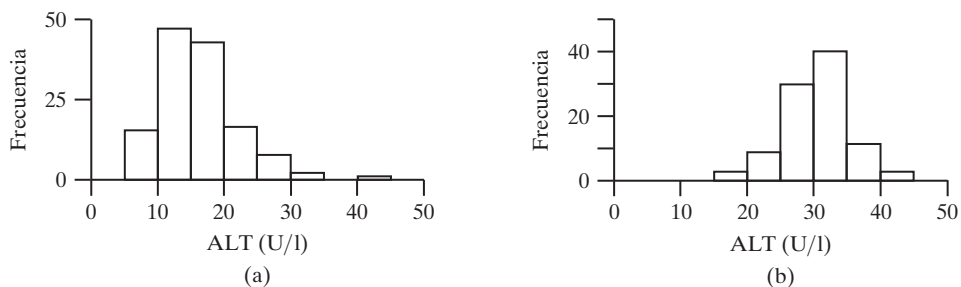


Figura 2.2.19 Distribución de medidas de ALT en suero: (a) de 129 voluntarios; (b) de 109 ensayos sobre el mismo espécimen

Ejercicios 2.2.1-2.2.9

2.2.1 Un paleontólogo midió la anchura (en mm) del último molar superior de 36 especímenes del mamífero extinto *Acropitecus rigidus*. Los resultados fueron los siguientes¹²:

6,1	5,7	6,0	6,5	6,0	5,7
6,1	5,8	5,9	6,1	6,2	6,0
6,3	6,2	6,1	6,2	6,0	5,7
6,2	5,8	5,7	6,3	6,2	5,7
6,2	6,1	5,9	6,5	5,4	6,7
5,9	6,1	5,9	5,9	6,1	6,1

- (a) Construya una distribución de frecuencias y preséntela en forma de tabla y de histograma.
(b) Describa la forma de la distribución.

2.2.2 En un estudio sobre la esquizofrenia, unos investigadores midieron la actividad de la enzima monoamina oxidasa (MAO) en las plaquetas de la sangre de 18 pacientes. Los resultados (expresados en nmoles de producto benzilaldehído por 108 plaquetas) fueron los siguientes¹³:

6,8	8,4	8,7	11,9	14,2	18,8
9,9	4,1	9,7	12,7	5,2	7,8
7,8	7,4	7,3	10,6	14,5	10,7

Construya un diagrama de puntos de los datos.

2.2.3 Considere los datos presentados en el Ejercicio 2.2.2. Construya una distribución de frecuencias y preséntela en forma de tabla y de histograma.

2.2.4 Un árbol dendrítico es una estructura ramificada que emana del cuerpo de una célula nerviosa. Como parte de un estudio sobre el desarrollo cerebral, se extrajeron 36 células nerviosas de cerebros de cerdos de Guinea recién nacidos. Los investigadores contaron el número de ramas dendríticas que emanaban de cada célula nerviosa. Los resultados fueron los siguientes¹⁴:

23	30	54	28	31	29	34	35	30
27	21	43	51	35	51	49	35	24
26	29	21	29	37	27	28	33	33
23	37	27	40	48	41	20	30	57

Construya un diagrama de puntos de los datos.

2.2.5 Considere los datos presentados en el Ejercicio 2.2.4. Construya una distribución de frecuencias y preséntela en forma de tabla y de histograma.

2.2.6 La cantidad total de proteína producida por una vaca lechera se puede estimar analizando su leche de forma periódica. A continuación siguen los valores de producción anual total de proteína (lb) de 28 vacas Holstein de dos años de edad. La dieta, los procedimientos de ordeño y otras condiciones fueron las mismas para todos los animales¹⁵:

425	481	477	434	410	397	438
545	528	496	502	529	500	465
539	408	513	496	477	445	546
471	495	445	565	499	508	426

Construya una distribución de frecuencias y preséntela en forma de tabla y de histograma.

2.2.7 Un veterinario midió la concentración de glucosa en la cámara anterior del ojo derecho y en el suero sanguíneo de 31 perros. Los datos que siguen son las medidas de glucosa de la cámara anterior, expresadas como porcentaje de la glucosa en sangre¹⁶:

81	85	93	93	99	76	75	84
78	84	81	82	89	81	96	82
74	70	84	86	80	70	131	75
88	102	115	89	82	79	106	

Construya una distribución de frecuencias y preséntela en forma de tabla y de histograma.

2.2.8 Unos agrónomos midieron el rendimiento de una variedad híbrida de maíz en 16 lugares de Illinois. Los datos en celemines por acre fueron¹⁷:

241	230	207	219	266	167
204	144	178	158	153	
187	181	196	149	183	

- (a) Construya un diagrama de puntos de los datos.
(b) Describa la forma de la distribución.

2.2.9 (Problema con computador) Los tripanosomas son parásitos que causan enfermedades en humanos y animales. En un estudio preliminar sobre la morfología de los tripanosomas, unos investigadores midieron las longitudes de 500 tripanosomas extraídos de la sangre de una rata. Los resultados se resumen en la distribución de frecuencias siguiente¹⁸.

Longitud (μm)	Frecuencia (número de individuos)	Longitud (μm)	Frecuencia (número de individuos)
15	1	27	36
16	3	28	41
17	21	29	48
18	27	30	28
19	23	31	43
20	15	32	27
21	10	33	23
22	15	34	10
23	19	35	4
24	21	36	5
25	34	37	1
26	44	38	1

- (a) Construya un histograma de los datos utilizando 24 clases (es decir, una clase para cada longitud entera, desde 15 hasta 38).
- (b) ¿Qué característica del histograma sugiere la interpretación de que los 500 individuos son una mezcla de dos tipos distintos?
- (c) Construya un histograma de los datos utilizando solo 6 clases. Comente cómo este histograma produce una impresión cualitativamente diferente a la del histograma del apartado (a).

2.3 Estadística descriptiva: medidas de tendencia central

En el caso de datos categóricos, la distribución de frecuencias proporciona un resumen conciso y completo de una muestra. En el caso de variables numéricas, la distribución de frecuencias se puede complementar de forma útil con algunas medidas numéricas. Una medida numérica calculada a partir de los datos de una muestra se denomina **estadístico***. Los **estadísticos descriptivos** son estadísticos que describen un conjunto de datos. Generalmente, los estadísticos descriptivos de una muestra se calculan para proporcionar información sobre una población de interés (véase la Sección 2.8). En esta sección estudiaremos medidas de tendencia central de los datos. Existen varias formas diferentes de definir el «centro» o «valor típico» de las observaciones de una muestra. Consideraremos las dos medidas de tendencia central más ampliamente utilizadas: la mediana y la media.

La mediana

Quizá la medida más simple del centro de un conjunto de datos sea la **mediana** muestral. La mediana muestral es el valor que está más cerca de la mitad de la muestra; es el valor de los datos que divide a dichos datos ordenados en dos mitades iguales. Para obtener la mediana, se ordenan primero las observaciones en orden creciente. En la secuencia de observaciones ordenadas, la mediana es el valor central (si n es impar) o el punto medio entre los dos valores del centro (si n es par). Denotaremos la mediana de la muestra mediante el símbolo $\tilde{y}$ (léase «y tilde»). El Ejemplo 2.3.1 ilustra estas definiciones.

Ejemplo 2.3.1 Ganancia de peso de corderos

Se presentan a continuación las ganancias de peso (lb) durante dos semanas de seis corderos jóvenes de la misma raza que habían seguido la misma dieta¹⁹:

Las observaciones ordenadas son

11 13 19 2 10 1

1 2 10 11 13 19

* Las medidas numéricas basadas en la población completa se denominan **parámetros**, que se comentan con más detalle en la Sección 2.8.

La mediana de la ganancia de peso es

$$\tilde{y} = \frac{10 + 11}{2} = 10,5 \text{ lb}$$

La mediana divide los datos ordenados en dos partes iguales (con el mismo número de observaciones por encima y por debajo de la mediana). La Figura 2.3.1 muestra un diagrama de puntos de los datos de ganancia de peso de corderos, junto con la posición de $\tilde{y}$. ■

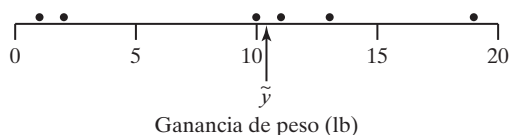


Figura 2.3.1 Diagrama de los datos de la ganancia de peso en corderos

Ejemplo 2.3.2 Ganancia de peso de corderos

Supongamos que la muestra contuviera un cordero más, y las siete observaciones ordenadas fueran como sigue:

1 2 10 10 11 13 19

Para esta muestra, la mediana de la ganancia de peso es

$$\tilde{y} = 10 \text{ lb}$$

(Nótese que en este ejemplo hay dos corderos cuya ganancia de peso es igual a la mediana. La cuarta observación, es decir, el segundo 10, es la mediana). ■

Una manera más formal de definir la mediana es en función de la posición en la secuencia ordenada (indicando la observación más pequeña como posición 1, la siguiente como 2, y así sucesivamente). La posición de la mediana es igual a

$$(0,5)(n + 1)$$

Es decir, si $n = 7$, calculamos $(0,5)(n + 1) = 4$, de forma que la mediana es la cuarta observación empezando a contar desde la más pequeña. Si $n = 6$, tenemos que $(0,5)(n + 1) = 3,5$, por lo que la mediana está en la mitad de las observaciones tercera y cuarta empezando a contar desde la más pequeña. Nótese que la fórmula $(0,5)(n + 1)$ no da el valor de la mediana, sino su posición en la lista ordenada de datos.

La media

La medida más familiar de tendencia central es el promedio ordinario o **media** (denominada algunas veces media aritmética). La media de una muestra (o «media muestral») es la suma de las observaciones dividida por el número de observaciones. Si denotamos por Y a una variable, entonces denotaremos las observaciones de la muestra por $y_1, y_2, \dots, y_n$, e indicaremos la media de la muestra como $\bar{y}$ (leído «y barra»). El Ejemplo 2.3.3 ilustra esta notación.

Ejemplo 2.3.3 Ganancia de peso de corderos

A continuación se muestran los datos del Ejemplo 2.3.1:

11 13 19 2 10 1

En este caso, $y_1 = 11$, y así sucesivamente hasta $y_6 = 1$. La suma de las observaciones es $11 + 13 + \dots + 1 = 56$. Podemos expresar esa suma utilizando la «notación de sumatorio», como $\sum_{i=1}^n y_i = 56$. El símbolo $\sum_{i=1}^n y_i$ significa «sumar los y_i ». Así, cuando $n = 6$, $\sum_{i=1}^n y_i = y_1 + y_2 + y_3 + y_4 + y_5 + y_6$. En este caso tenemos $\sum_{i=1}^n y_i = 11 + 13 + 19 + 2 + 10 + 1 = 56$.

La ganancia media de peso de los seis corderos de esta muestra es

$$\bar{y} = \frac{11 + 13 + 19 + 2 + 10 + 1}{6} = \frac{56}{6} = 9,33 \text{ lb}$$

La media muestral La definición general de media muestral es

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

donde los valores y_i son las observaciones de la muestra y n es el tamaño de la muestra (es decir, el número de y_i).

Mientras que la mediana divide los datos en dos partes iguales (es decir, el mismo número de observaciones por encima y por debajo), la media es el «punto de equilibrio» de los datos. La Figura 2.3.2 muestra el diagrama de puntos de los datos de ganancia de peso en corderos, junto con la posición de $\bar{y}$. Si los puntos de datos fueran niños en un columpio sin peso, entonces el columpio se inclinaría si el punto de apoyo se situara en $\bar{y}$, independientemente de que haya el mismo número de niños en cada lado. Los niños del lado izquierdo (por debajo de $\bar{y}$) se sientan más lejos de $\bar{y}$ que los niños de la derecha (por encima de $\bar{y}$), por lo que el columpio se inclina. Sin embargo, si el punto de apoyo se sitúa en $\bar{y}$, el columpio quedaría en equilibrio, como muestra la Figura 2.3.3. ■

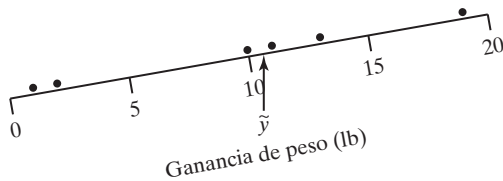


Figura 2.3.2 Diagrama de los datos de la ganancia de peso en corderos con la mediana muestral como el punto de apoyo de la balanza

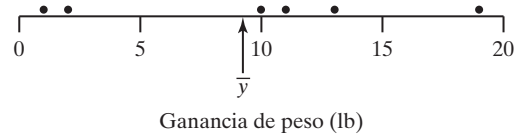


Figura 2.3.3 Diagrama de los datos de la ganancia de peso en corderos con la media muestral como el punto de apoyo de la balanza

La diferencia entre un punto de datos y la media se denomina **desviación**: $\text{desviación}_i = y_i - \bar{y}$. La media tiene la propiedad de que la suma de las desviaciones respecto a la media es cero, es decir $\sum_{i=1}^n (y_i - \bar{y}) = 0$. En este sentido, la media es el centro de la distribución: las desviaciones positivas equilibran a las desviaciones negativas.

Ejemplo 2.3.4 Ganancia de peso de corderos

En el ejemplo de la ganancia de peso de los corderos, las desviaciones son las siguientes:

$$\text{desviación}_1 = y_1 - \bar{y} = 11 - 9,33 = 1,67$$

$$\text{desviación}_2 = y_2 - \bar{y} = 13 - 9,33 = 3,67$$

$$\text{desviación}_3 = y_3 - \bar{y} = 19 - 9,33 = 9,67$$

$$\text{desviación}_4 = y_4 - \bar{y} = 2 - 9,33 = -7,33$$

$$\text{desviación}_5 = y_5 - \bar{y} = 10 - 9,33 = 0,67$$

$$\text{desviación}_6 = y_6 - \bar{y} = 1 - 9,33 = -8,33$$

La suma de las desviaciones es $\sum_{i=1}^n (y_i - \bar{y}) = 1,67 + 3,67 + 9,67 - 7,33 + 0,67 - 8,33 = 0$. ■

Robustez Se dice que un estadístico es **robusto** o **resistente** si el valor de dicho estadístico es afectado relativamente poco por cambios en una parte pequeña de los datos, incluso si dichos cambios son importantes. La mediana es un estadístico robusto, pero la media no lo es, ya que se puede desplazar mucho por cambios incluso en una sola observación. El Ejemplo 2.3.5 ilustra este comportamiento.

Ejemplo 2.3.5 Ganancia de peso de corderos

Recuérdese que para los datos de ganancia de peso de corderos

1 2 10 11 13 19

obtuvimos

$$\bar{y} = 9,3 \quad \text{y} \quad \tilde{y} = 10,5$$

Supongamos ahora que la observación 19 se cambia, o incluso se omite. ¿Cómo quedarían afectadas la media y la mediana? Podemos visualizar el efecto si imaginamos que el punto más a la derecha de la Figura 2.3.3 se mueve o se elimina. Claramente, la media puede cambiar bastante, mientras que la mediana en general será menos afectada. Por ejemplo:

- Si el 19 se cambia por 12, la media resulta ser 8,2 y la mediana no cambia.
- Si el 19 se elimina, la media pasa a valer 7,4 y la mediana pasa a valer 10.

Estos cambios no son exagerados, es decir, podrían haber surgido a partir del mismo experimento. Por supuesto, un cambio enorme, como cambiar el 19 por 100, desplazaría drásticamente la media. Nótese que, sin embargo, la mediana no cambiaría. ■

Visualización de la media y la mediana

Podemos visualizar la media y la mediana en relación con el histograma de una distribución. La mediana divide el área bajo el histograma aproximadamente por la mitad, ya que divide a las observaciones aproximadamente por la mitad [«aproximadamente», porque varias observaciones pueden coincidir con la mediana, como en el Ejemplo 2.3.3(b), y porque las observaciones dentro de cada clase pueden no estar uniformemente distribuidas dentro de la clase]. La media se puede visualizar como el punto de equilibrio del histograma. Si el histograma estuviera hecho de contrachapado, estaría en equilibrio si se apoyara en la media.

Si la distribución de frecuencias es simétrica, la media y la mediana son iguales y están situadas en el centro de la distribución. Si la distribución de frecuencias está sesgada, ambas medidas se desplazan hacia la cola más larga, pero la media en general se desplaza más que la mediana. El efecto de sesgo se ilustra mediante el ejemplo siguiente.

Ejemplo 2.3.6 Tiempos de canto de grillos

Los grillos Mormon machos (*Anabrus simplex*) cantan para aparearse. Un investigador de campo midió la duración de 51 cantos sin éxito, es decir, el tiempo transcurrido hasta que el macho que cantaba renunciaba y abandonaba su posición²⁰. La Figura 2.3.4 muestra el histograma de los 51 tiempos de canto. La Tabla 2.3.1 muestra todos los datos. La mediana es 3,7 min y la media es 4,3 min. La discrepancia entre estas medidas es debida en gran parte a la cola larga e irregular de la distribución. Los pocos tiempos de canto inusualmente largos influyen en la media, pero no en la mediana. ■

Tabla 2.3.1 51 tiempos de canto de grillos (min)

4,3	3,9	17,4	2,3	0,8	1,5	0,7	3,7
24,1	9,4	5,6	3,7	5,2	3,9	4,2	3,5
6,6	6,2	2,0	0,8	2,0	3,7	4,7	
7,3	1,6	3,8	0,5	0,7	4,5	2,2	
4,0	6,5	1,2	4,5	1,7	1,8	1,4	
2,6	0,2	0,7	11,5	5,0	1,2	14,1	
4,0	2,7	1,6	3,5	2,8	0,7	8,6	

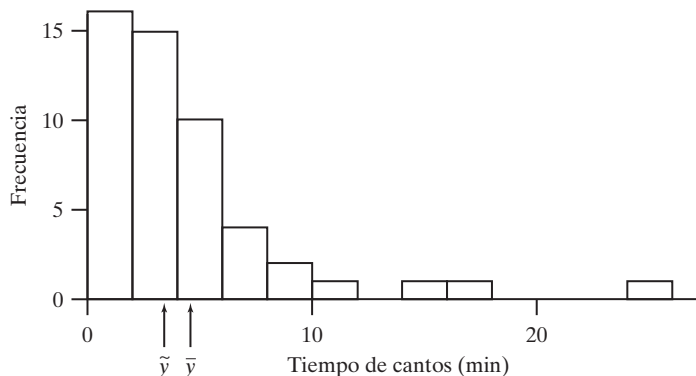


Figura 2.3.4 Histograma de los tiempos de canto de grillos

Media versus mediana

Tanto la media como la mediana son en general medidas razonables de la tendencia central de un conjunto de datos. La media está relacionada con la suma. Por ejemplo, si la media de la ganancia de peso de 100 corderos es de 9 lb, entonces la ganancia total de peso fue de 900 lb, y este total puede ser de interés si se traduce más o menos directamente en beneficio para el granjero. En algunas situaciones la media tiene muy poco sentido. Supongamos, por ejemplo, que las observaciones son tiempos de supervivencia de pacientes de cáncer con un cierto protocolo de tratamiento, y que la mayoría de los pacientes sobrevive menos de un año, mientras que unos pocos responden bien y sobreviven 5 o incluso 10 años. En este caso, el tiempo medio de supervivencia sería mayor que el tiempo de supervivencia de la mayoría de los pacientes. La mediana representaría de forma más precisa la experiencia de un paciente «típico». Nótese también que el tiempo medio de supervivencia no se puede calcular hasta que el último paciente ha fallecido. La mediana no tiene esta desventaja. Situaciones en las que la mediana se puede calcular fácilmente pero la media no, no son raras en bioensayos, y en estudios de supervivencia y de toxicidad.

Hemos visto que la mediana es más resistente que la media. Si un conjunto de datos contiene unas pocas observaciones que están lejos del cuerpo principal de los datos (es decir, una cola larga e «irregular»), entonces la media estará excesivamente influida por esas pocas observaciones inusuales. Entonces la «cola» puede «mover al perro», una situación no deseable. En esos casos, la resistencia de la mediana puede ser ventajosa.

Una ventaja de la media es que en algunas circunstancias es más eficiente que la mediana. La eficiencia es una noción técnica de teoría estadística. De forma simple, un método es eficiente si aprovecha completamente toda la información de los datos. En parte debido a esta eficiencia, la media ha tenido un papel principal en los métodos clásicos de estadística.

Ejercicios 2.3.1-2.3.16

2.3.1 Invente una muestra de tamaño 5 tal que su media muestral sea 20 y tal que no todas las observaciones sean iguales.

2.3.2 Invente una muestra de tamaño 5 tal que su media muestral sea 20 y su mediana muestral sea 15.

2.3.3 Un investigador aplicó el componente carcinógeno (causante de cáncer) benzo(a)pireno en la piel de cinco ratones, y midió la concentración en el tejido del hígado después de 48 horas. Los resultados (nmol/g) fueron como sigue²¹:

6,3 5,9 7,0 6,9 5,9

Determine la media y la mediana.

2.3.4 Considere los datos del Ejercicio 2.3.3. ¿Permiten la media y la mediana calculadas deducir que, en general, la concentración en el tejido del hígado después de 48 horas es diferente de 6,3 nmol/g?

2.3.5 Seis hombres con colesterol alto en el suero participaron en un estudio para evaluar los efectos de la dieta en los niveles de colesterol. Al comenzar el estudio sus niveles de colesterol en suero (mg/dl) eran los siguientes²²:

366 327 274 292 274 230

Determine la media y la mediana.

2.3.6 Considere los datos del Ejercicio 2.3.5. Suponga que se añade a la muestra una observación adicional igual a 400. ¿Cuáles serían ahora la media y la mediana de las siete observaciones?

2.3.7 Se midió la ganancia de peso de siete novillos durante un periodo de prueba de 140 días. Las ganancias prome-

dio diarias (lb/día) de nueve novillos con la misma dieta fueron las siguientes²³:

3,89 3,51 3,97 3,31 3,21
3,36 3,67 3,24 3,27

Determine la media y la mediana.

2.3.8 Considere los datos del Ejercicio 2.3.7. ¿Permiten la media y la mediana calculadas deducir que, en general, los novillos ganaron 3,5 lb/día? ¿Son consistentes con una ganancia de 4 lb/día?

2.3.9 Considere los datos del Ejercicio 2.3.7. Suponga una observación adicional de valor 2,46 que se añade a la muestra. ¿Cuáles serían la media y la mediana de las 10 observaciones?

2.3.10 Como parte de un experimento clásico sobre mutaciones, se tomaron 10 conjuntos de idéntico tamaño del mismo cultivo de la bacteria *E. coli*. En cada conjunto se determinó el número de bacterias resistentes a un cierto virus. Los resultados fueron los siguientes²⁴:

14 15 13 21 15
14 26 16 20 13

(a) Construya la distribución de frecuencias de estos datos y presente la forma del histograma.

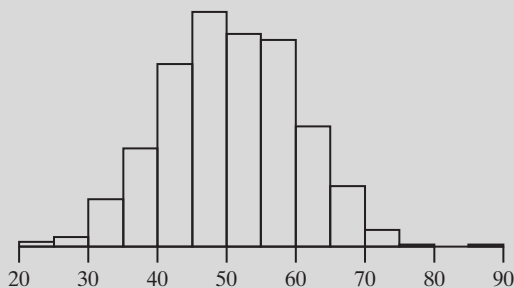
(b) Determine la mediana de los datos y marque sus posiciones en el histograma.

2.3.11 La tabla que sigue muestra el tamaño de la camada (número de cochinillos supervivientes después de 21 días) en 36 cerdas (como en el Ejemplo 2.2.4). Determine la mediana de la camada. (Sugerencia: nótese que solo hay un 5, pero hay dos 7, tres 8, etc.).

Número de cochinitos	Frecuencia (número de cerdas)
5	1
6	0
7	2
8	3
9	3
10	9
11	8
12	5
13	3
14	2
Total	36

2.3.12 Considere los datos del Ejercicio 2.3.11. Determine la media de las 36 observaciones. (*Sugerencia:* nótese que hay un 5 pero hay dos 7, tres 8, etc. Por tanto, $\sum y_i = 5 + 7 + 7 + 8 + 8 + 8 + \dots = 5 + 2(7) + 3(8) + \dots$).

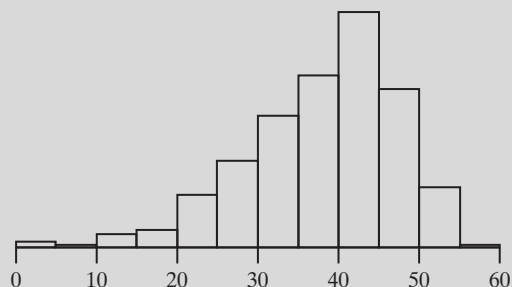
2.3.13 Se presenta a continuación un histograma.



- (a) Estime la mediana de la distribución.
(b) Estime la media de la distribución.

2.3.14 Considere el histograma del Ejercicio 2.3.13. «Leyendo» el histograma, estime el porcentaje de observaciones que son menores de 40. ¿Está este porcentaje cerca del 15 %, el 25 %, el 35 % o el 45 %? (*Nota:* no se da la escala de frecuencias en este histograma, porque no hay necesidad de calcular el número de observaciones de cada clase. El porcentaje de observaciones que son menores de 40 se puede estimar observando el área).

2.3.15 Se presenta a continuación un histograma.



- (a) Estime la mediana de la distribución.
(b) Estime la media de la distribución.

2.3.16 Considere el histograma del Ejercicio 2.3.15. «Leyendo» el histograma, estime el porcentaje de observaciones que son mayores de 45. ¿Está este porcentaje cerca del 15 %, el 25 %, el 35 % o el 45 %? (*Nota:* no se da la escala de frecuencias en este histograma, porque no hay necesidad de calcular el número de observaciones de cada clase. El porcentaje de observaciones que son mayores de 45 se puede estimar observando el área).

2.4 Diagramas de caja

Uno de los gráficos más eficientes, tanto para examinar una única distribución como para hacer comparaciones entre distribuciones, se conoce como diagrama de caja, de lo que trata esta sección. Antes de presentar los diagramas de caja, sin embargo, necesitamos hablar de los cuartiles.

Cuartiles e intervalo intercuartílico

La mediana de la distribución divide dicha distribución en dos partes, una parte inferior de la parte superior. Los **cuartiles** de una distribución dividen cada una de esas partes por la mitad, dividiendo por tanto la distribución en cuatro cuartos. El **primer cuartil**, denominado Q_1 , es la mediana de los valores de los datos de la mitad inferior del conjunto de

datos. El **tercer cuartil**, denominado Q_3 , es la mediana de los valores de los datos de la mitad superior del conjunto de datos*. El siguiente ejemplo ilustra estas definiciones.

Ejemplo 2.4.1 Presión sanguínea

Las presiones sistólicas de la sangre (mmHg) de siete hombres de mediana edad fueron las siguientes²⁵:

151 124 132 170 146 124 113

Ordenando estos valores, la muestra es

113 124 124 132 146 151 170

La mediana es la cuarta observación más grande, que es 132. Hay tres puntos de datos en la parte inferior de la distribución: 113, 124 y 124. La mediana de esos tres valores es 124. Por tanto el primer cuartil, Q_1 , es 124.

Asimismo, hay tres puntos de datos en la parte superior de la distribución: 146, 151 y 170. La mediana de estos tres valores es 151. Por tanto, el tercer cuartil, Q_3 , es 151.

113	124	124	132	146	151	170
	↑		⋮		↑	
	primer cuartil		mediana		tercer cuartil	
	Q_1				Q_3	

Nótese que la mediana no está incluida ni en la parte inferior ni en la parte superior de la distribución. Si el tamaño de la muestra, n , es par, entonces exactamente la mitad de las observaciones estarán en la parte inferior de la distribución y la otra mitad estarán en la parte superior.

El **intervalo intercuartílico** es la diferencia entre los cuartiles tercero y primero, y se abrevia como **IQR**: $IQR = Q_3 - Q_1$. Para el caso de los datos de presión sanguínea del Ejemplo 2.4.1, el IQR es $151 - 124 = 27$.

Ejemplo 2.4.2 Pulso

Se midió el pulso a 12 estudiantes universitarios²⁶. Se presentan a continuación los datos, dispuestos en orden creciente, con la posición de la mediana indicada por la línea de puntos:

62 64 68 70 70 74 ⋮ 74 76 76 78 78 80

La mediana es $\frac{74 + 74}{2} = 74$. Hay seis observaciones en la parte inferior de la distribución: 62, 64, 68, 70, 70, 74.

Por tanto, el primer cuartil es el promedio de los valores tercero y cuarto de los datos ordenados:

$$Q_1 = \frac{68 + 70}{2} = 69$$

Hay seis observaciones en la parte superior de la distribución: 74, 76, 76, 78, 78, 80. Por tanto, el tercer cuartil es el promedio de los valores ordenados noveno y décimo (los valores ordenados tercero y cuarto de la parte superior de la distribución):

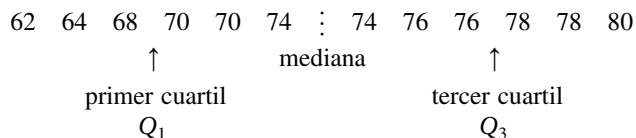
$$Q_3 = \frac{76 + 78}{2} = 77$$

Por tanto, el intervalo intercuartílico es

$$IQR = 77 - 69 = 8$$

* Algunos autores y también algún *software* de computador utilizan otras definiciones de cuartiles. Una definición alternativa común es decir que el primer cuartil ocupa la posición $(.25)(n + 1)$. Por tanto, si $n = 10$, el primer cuartil estaría en la posición $(.25)(11) = 2.75$. Es decir, para obtener el primer cuartil tendríamos que interpolar entre las observaciones ordenadas segunda y tercera. Si n es grande, entonces existen pocas diferencias prácticas entre las definiciones que utilizan los diversos autores.

Tenemos entonces

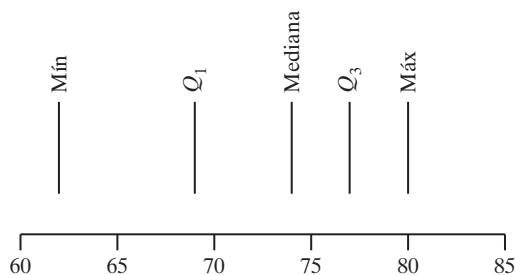


El mínimo valor del pulso es 62 y el máximo es 80. ■

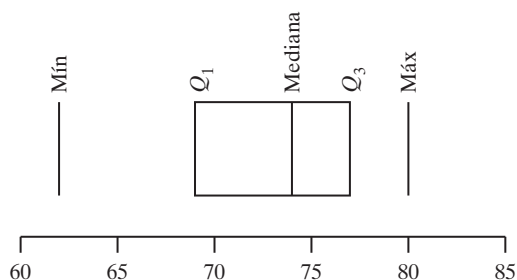
El mínimo, el máximo, la mediana y los cuartiles, considerados juntos, se denominan **resumen de cinco números** de los datos.

Diagramas de caja

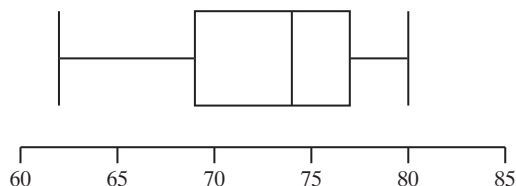
Un **diagrama de caja** es una representación visual del resumen de cinco números. Para hacer un diagrama de caja se representa en primer lugar una recta numérica. Seguidamente se marcan sobre dicha recta las posiciones del mínimo, Q_1 , la mediana, Q_3 , y el máximo:



Seguidamente, se construye una caja conectando los cuartiles:



Nótese que el intervalo intercuartílico es igual a la longitud de la caja. Finalmente, extendemos unas «patillas» desde Q_1 hacia abajo hasta el mínimo y desde Q_3 hacia arriba hasta el máximo:



Un diagrama de caja proporciona un resumen visual rápido de la distribución. Podemos ver inmediatamente dónde está el centro de los datos observando la línea dentro de la caja que localiza a la mediana. Podemos ver la extensión de la

La gente utiliza el término «*outlier*» informalmente. Hay, sin embargo, una definición común de «*outlier*» en la práctica estadística. Para dar una definición de *outlier*, presentaremos primero lo que se conoce como valla. La **valla inferior** de la distribución es

$$\text{valla inferior} = Q_1 - 1,5 \times \text{IQR}$$

La **valla superior** de una distribución es

$$\text{valla superior} = Q_3 + 1,5 \times \text{IQR}$$

Esto significa que las vallas están situadas a 1,5 IQRs (es decir, 1,5 la longitud de la caja) más allá del límite de la caja en un diagrama de caja.

Nótese que las vallas no tienen por qué ser valores de datos. De hecho, podría no haber datos cerca de las vallas. Las vallas simplemente sitúan límites dentro de la distribución de las muestras. Estos límites nos proporcionan una manera de definir los *outliers*. *Un outlier es un punto de datos que está fuera de las vallas*. Es decir, si

$$\text{punto de datos} < Q_1 - 1,5 \times \text{IQR}$$

o

$$\text{punto de datos} > Q_3 + 1,5 \times \text{IQR}$$

entonces diremos que el punto es un *outlier*.

Ejemplo 2.4.4 Pulso

En el Ejemplo 2.4.2 vimos que $Q_1 = 69$, $Q_3 = 77$ y $\text{IQR} = 8$. Por tanto, la valla inferior es $69 - 1,5 \times 8 = 69 - 12 = 57$. Todo punto menor que 57 sería un *outlier*. La valla superior es $77 + 1,5 \times 8 = 77 + 12 = 89$. Todo punto mayor que 89 sería un *outlier*. Como no hay puntos menores que 57 ni mayores que 89, en este conjunto de datos no hay *outliers*. ■

Ejemplo 2.4.5 Crecimiento de rábanos con luz

Los datos del Ejemplo 2.4.3 correspondían a semillas de rábanos que crecían en total oscuridad. En otra parte del experimento los estudiantes hicieron crecer 14 semillas de rábanos con luz constante. Las observaciones, ya ordenadas, son

3	5	5	7	7	8	9	10	10	10	10	14	20	21
			↑				↑				↑		
			primer cuartil			mediana					tercer cuartil		
			Q_1								Q_3		

Por tanto, la mediana es $\frac{9 + 10}{2} = 9,5$, Q_1 es 7, y Q_3 es 10. El intervalo intercuartílico es $\text{IQR} = 10 - 7 = 3$. La valla inferior es $7 - 1,5 \times 3 = 7 - 4,5 = 2,5$, por lo que todo punto menor que 2,5 sería un *outlier*. La valla superior es $10 + 1,5 \times 3 = 10 + 4,5 = 14,5$. Todo punto mayor que 14,5 es un *outlier*. Por tanto, las dos observaciones mayores de este conjunto de datos son *outliers*: 20 y 21. ■

El método que hemos definido para identificar *outliers* permite que sea la mayor parte de los datos la que determine cuán extrema debe ser una observación antes de considerarla un *outlier*, ya que los cuartiles y el IQR están determinados por los propios datos. Por tanto, un punto que sea un *outlier* en un conjunto de datos podría no serlo en otro conjunto de datos. Etiquetaremos un punto como *outlier* si tiene un valor inusual respecto a la variabilidad inherente de todo el conjunto de datos.

Después de que un *outlier* ha sido identificado, a menudo la gente está tentada de eliminarlo del conjunto de datos. En general, esto no es una buena idea. Si podemos identificar que un *outlier* se ha producido debido, por ejemplo, a un error de equipamiento, entonces tenemos una buena razón para eliminar el *outlier* antes de analizar el resto de los datos. Sin embargo, ocurre muy a menudo que los *outliers* aparecen en los conjuntos de datos sin ninguna razón externa identificable para los mismos. En esos casos, simplemente procederemos con nuestro análisis, teniendo en cuenta que hay un

outlier presente. En algunos casos, podríamos desear calcular la media, por ejemplo, con y sin el *outlier*, y presentar ambos resultados, para mostrar el efecto del *outlier* en el análisis global. Esto es preferible a eliminar el *outlier*, que oscurecería el hecho de que estaba presente un dato inusual. Al presentar los datos gráficamente, podemos llamar la atención sobre los *outliers* utilizando diagramas de caja modificados, que presentamos a continuación.

Diagramas de caja modificados

Una variación estándar de la idea de diagrama de caja es lo que se conoce como diagrama de caja modificado. Un **diagrama de caja** modificado es un diagrama de caja en el que los *outliers*, si existen, se representan como puntos separados. La ventaja de un diagrama de caja modificado es que nos permite ver rápidamente donde están los *outliers*, si existen.

Para hacer un diagrama de caja modificado, procederemos como hicimos para hacer un diagrama de caja, excepto en el último paso. Tras dibujar la caja del diagrama de caja, comprobaremos si existen *outliers*. Si no hay *outliers*, entonces llevaremos las patillas de la caja hasta los extremos (el mínimo y el máximo). Sin embargo, si hay *outliers* en la parte superior de la distribución, los identificaremos mediante un punto u otro símbolo. Después extenderemos la patilla desde Q_3 hasta el punto de datos de mayor valor que no sea un *outlier*. Asimismo, si hay *outliers* en la parte inferior de la distribución, los identificaremos con asteriscos y extenderemos una patilla desde Q_1 hasta la observación de menor valor que no sea un *outlier*. La Figura 2.4.2 muestra la distribución del crecimiento de semillas de rábanos bajo luz constante. El área entre las vallas superior e inferior es blanca mientras que la región de los *outliers* es sombreada.

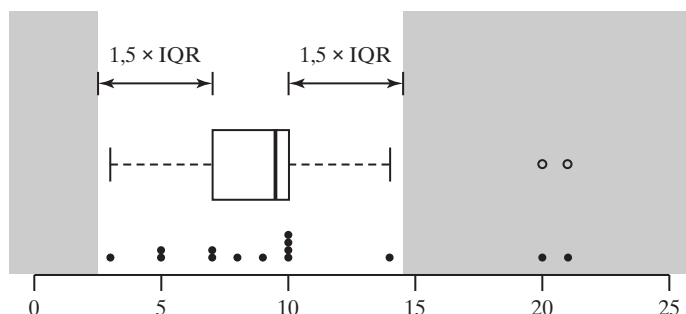


Figura 2.4.2 Diagrama de puntos y diagrama de caja de los datos sobre crecimiento de rábanos en luz constante. Los puntos en la región sombreada son *outliers*

La Figura 2.4.3 muestra un diagrama de caja y diagrama de caja modificado de los datos de crecimiento de semillas de rábanos en luz constante.

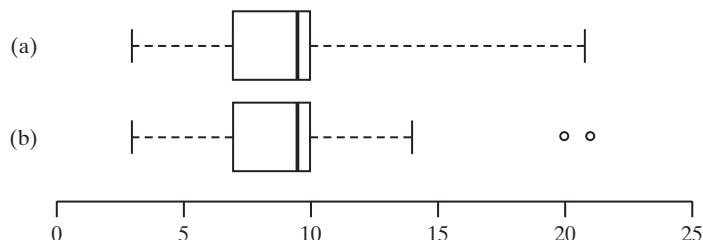


Figura 2.4.3 (a) diagrama de caja de los datos sobre crecimiento de rábanos en luz constante; (b) diagrama de caja modificado en los datos sobre crecimiento de rábanos

Muy frecuentemente, cuando la gente hace diagramas de caja, en realidad realiza diagramas de caja modificados. El *software* estadístico está en general programado para producir un diagrama de caja modificado cuando el usuario pide un diagrama de caja. Por tanto, cuando utilizemos el término «diagrama de caja» queremos decir «diagrama de caja modificado».

Ejercicios 2.4.1-2.4.8

2.4.1 Se presentan a continuación los datos del Ejercicio 2.3.10 sobre el número de bacterias resistentes al virus en cada uno de los 10 conjuntos:

14	15	13	21	15
14	26	16	20	13

- Determine la mediana y los cuartiles.
- Determine el intervalo intercuartílico.
- ¿Qué valor debería tener una observación de estos datos puestos en orden para ser un *outlier*?

2.4.2 Se presentan a continuación las 18 medidas de la actividad MAO del Ejercicio 2.2.2:

6,8	8,4	8,7	11,9	14,2	18,8
9,9	4,1	9,7	12,7	5,2	7,8
7,8	7,4	7,3	10,6	14,5	10,7

- Determine la mediana y los cuartiles.
- Determine el intervalo intercuartílico.
- ¿Qué valor debería tener una observación de estos datos puestos en orden para ser un *outlier*?
- Construya un diagrama de caja (modificado) de los datos.

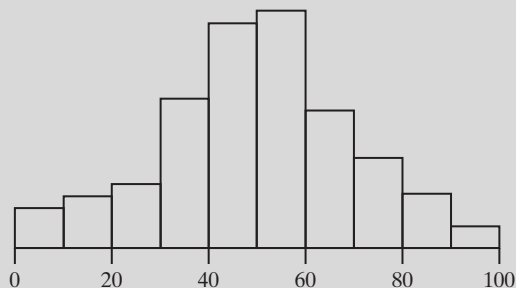
2.4.3 En un estudio sobre producción de leche de ovejas (para su uso en la fabricación de quesos) un investigador midió la leche producida durante tres meses de 11 ovejas. Las producciones (en litros) fueron las siguientes²⁸:

56,5	89,8	110,1	65,6	63,7	82,6
75,1	91,5	102,9	44,4	108,1	

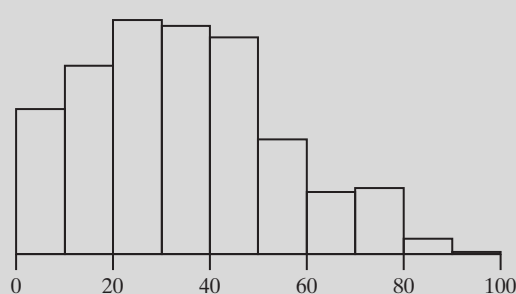
- Determine la mediana y los cuartiles.
- Determine el intervalo intercuartílico.
- Construya un diagrama de caja (modificado) de los datos.

2.4.4 En cada uno de los siguientes histogramas utilice dichos histogramas para estimar la media de los cuartiles. Construya después un diagrama de caja de la distribución.

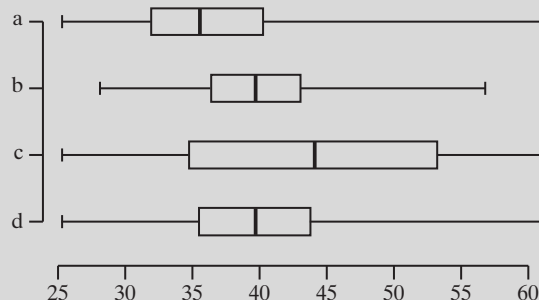
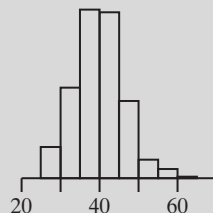
(a)



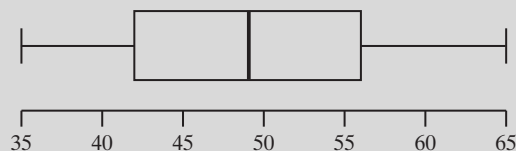
(b)



2.4.5 El presente histograma muestra los mismos datos correspondientes a uno de los cuatro diagramas de caja. ¿Qué diagrama de caja corresponde al histograma? Explique su respuesta.



2.4.6 El siguiente diagrama de caja muestra el resumen de cinco números de un conjunto de datos. El mínimo de dichos datos es 35, Q_1 es 42, la media es 49, Q_3 es 56 y el máximo es 65. ¿Es posible que no haya ninguna observación del conjunto de datos que sea igual a 42? Explique su respuesta.



2.4.7 Se puede utilizar *software* estadístico para obtener el resumen de cinco números de un conjunto de datos. Se presenta a continuación un ejemplo del resumen de estadística

descriptiva de MINITAB de una variable almacenada en la columna 1 (C1) de la hoja de cálculo de MINITAB.

Variable	N	Mean	Median	TrMean	StDev	SEMean
C1	75	119.94	118.40	119.98	9.98	1.15
Variable	Min	Max	Q1	Q3		
C1	95.16	145.11	113.59	127.42		

(a) Utilice la salida de MINITAB para calcular el intervalo intercuartílico.

(b) ¿Hay *outliers* en este conjunto de datos?

2.4.8 Considere los datos del Ejercicio 2.4.7. Utilice el resumen de cinco números que se proporciona para crear un diagrama de caja de los datos.

2.5 Relaciones entre variables

En la sección anterior hemos estudiado resúmenes **univariados** de variables tanto numéricas como categóricas. Un resumen univariado es un resumen gráfico o numérico de una sola variable.

El histograma, diagrama de caja, media muestral y mediana son todos ejemplos de resúmenes univariados de datos numéricos. El diagrama de barras y las tablas de frecuencias y de frecuencias relativas son ejemplos de resúmenes univariados de datos categóricos. En esta sección presentamos algunos resúmenes gráficos **bivariados** comunes que se usan para examinar la *relación* entre parejas de variables.

Relaciones categórica-categórica

Para entender la relación entre dos variables categóricas, resumiremos primero los datos en una **tabla de frecuencias bivariada**. A diferencia de la tabla de frecuencias presentada en la Sección 2.2 (una tabla univariada), la tabla de frecuencias bivariada tiene filas y columnas (una dimensión para cada variable). La selección de qué variable va en las filas y qué variable va en las columnas es arbitraria. El siguiente ejemplo considera la relación entre dos variables categóricas: fuente de *E. coli* y localización del muestreo.

Ejemplo 2.5.1 Contaminación de *E. coli* entre cuencas

En un esfuerzo para determinar si hay diferencias en las fuentes primarias de contaminación fecal en diferentes localizaciones de la cuenca de Morro Bay, se recogieron $n = 623$ especímenes de agua en tres posiciones primarias que pertenecían a la cuenca de Morro Bay: Chorro Creek ($n_1 = 241$), Los Osos Creek ($n_2 = 256$) y Baywood Seeps ($n_3 = 126$)²⁹. Se utilizaron técnicas de reconocimiento de ADN para determinar el origen intestinal de la variedad de *E. coli* dominante en cada espécimen de agua. Los orígenes de *E. coli* se clasificaron en las siguientes cinco categorías: pájaros, animales domésticos (por ejemplo, gatos o perros), animales de granja (por ejemplo, caballos, vacas o cerdos), humanos u otros mamíferos terrestres (por ejemplo, zorro, ratón o coyote). Por tanto, cada espécimen de agua tenía *dos* variables categóricas medidas: localización (Chorro, Los Osos o Baywood) y fuente de *E. coli* (pájaro, ..., mamífero terrestre). La Tabla 2.5.1 presenta una tabla de frecuencias de los datos. ■

Tabla 2.5.1 Tabla de frecuencias de fuentes de *E. coli* por localización

Localización	Fuente de <i>E. coli</i>					Total
	Pájaro	Animal doméstico	Animal de granja	Humano	Mamífero terrestre	
Chorro Creek	46	29	106	38	22	241
Los Osos Creek	79	56	32	63	26	256
Baywood	35	23	0	60	8	126
Total	160	108	138	161	56	623

Aunque la Tabla 2.5.1 proporciona un resumen conciso de los datos, es difícil descubrir patrones en los mismos. Examinar las frecuencias relativas (proporciones por filas o columnas) ayuda a menudo a realizar comparaciones útiles, como se puede ver en el siguiente ejemplo.

Ejemplo 2.5.2 Contaminación de *E. coli* entre cuencas

¿Son los animales domésticos un problema mayor (respecto a la fuente) de *E. coli* en Chorro Creek o en Baywood? La Tabla 2.5.1 muestra que el número de *E. coli* procedentes de animales domésticos es mayor en Chorro (29) que en Baywood (23), por lo que a primera vista parece que los animales domésticos son más problemáticos en Chorro. Sin embargo, como se recogieron más especímenes de agua en Chorro ($n_1 = 241$) que en Baywood ($n_2 = 126$), la frecuencia relativa de fuente de *E. coli* debida a animales domésticos es realmente menor en Chorro ($29/241 = 0,120$) que en Baywood ($23/126 = 0,183$). La Tabla 2.5.2 muestra los porcentajes por filas, facilitando así las comparaciones de las fuentes de *E. coli* entre localizaciones. (Nótese que los porcentajes por columnas no tendrían sentido en este contexto, ya que el agua fue muestreada por localización y no por fuente de *E. coli*). ■

Tabla 2.5.2 Tabla de frecuencias relativas bivariadas (porcentajes por filas) de fuentes de <i>E. coli</i> por localización						
Localización	Fuente de <i>E. coli</i>					Total
	Pájaro	Animal doméstico	Animal de granja	Humano	Mamífero terrestre	
Chorro Creek	19,1	12,0	44,0	15,8	9,1	100
Los Osos Creek	30,9	21,9	12,5	24,6	10,2	100
Baywood	27,8	18,3	0,0	47,6	6,3	100
Total	25,7	17,3	22,2	25,8	9,0	100

Para visualizar los datos de las Tablas 2.5.1 y 2.5.2 podemos examinar diagramas de barras apiladas. En un **diagrama de barras apiladas**, la altura total de cada barra refleja el tamaño de la muestra para un nivel de la variable categórica X (por ejemplo, la localización), y la altura o espesor de cada segmento de los que componen la barra representa el valor de la variable categórica Y (por ejemplo, la fuente de *E. coli*). La Figura 2.5.1 muestra un diagrama de barras apiladas para las distintas fuentes de *E. coli* en las distintas cuencas que se presentan en la Tabla 2.5.1.

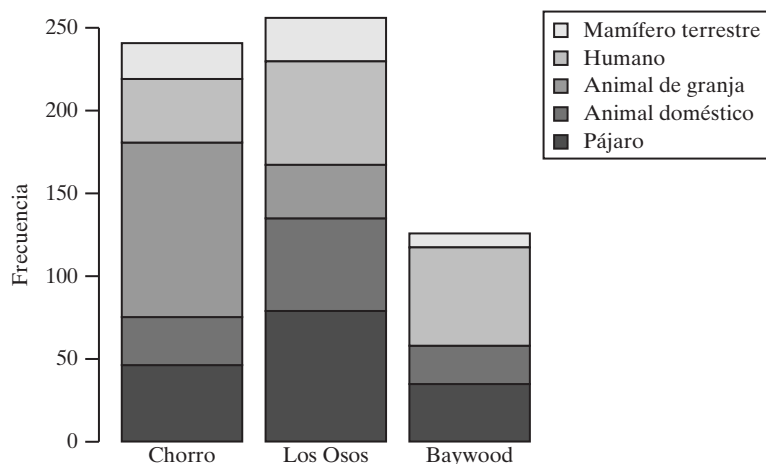


Figura 2.5.1 Diagrama de frecuencias apiladas de las fuentes de *E. coli* por localización

Como en el caso de la tabla de frecuencias, el diagrama de barras apiladas correspondiente a frecuencias no permite hacer comparaciones entre las tres localizaciones ya que el tamaño de la muestra es diferente en cada una de ellas. (Este tipo de gráfica ayuda a resaltar la diferencia en el tamaño de las muestras; por ejemplo, es muy claro que se tomaron muchos menos especímenes de agua en Baywood). Un diagrama que muestra mejor la distribución de una variable categórica con respecto a los niveles de otra es un **diagrama de barras apiladas de frecuencias relativas** (porcentajes), que presenta en forma gráfica los resúmenes de una tabla bivariada de frecuencias relativas como la Tabla 2.5.2. La Figura 2.5.2 muestra un ejemplo utilizando los datos de contaminación por *E. coli*. Este diagrama normaliza las barras de la Figura 2.5.1 de forma que tienen la misma altura (100 %), lo que facilita las comparaciones entre las tres localizaciones.

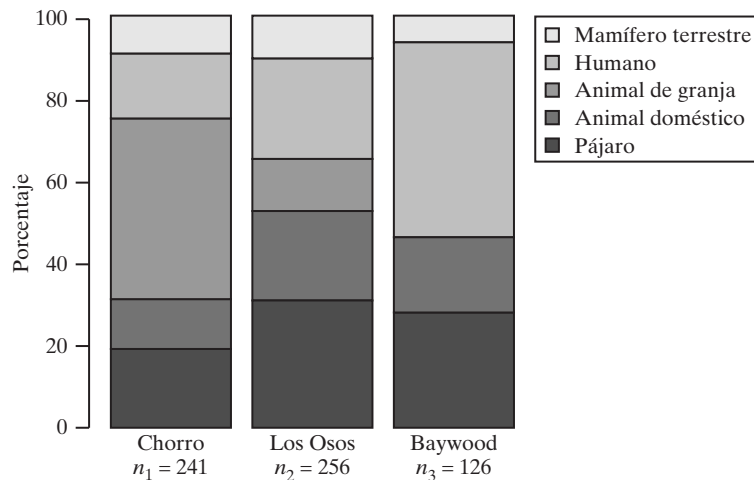


Figura 2.5.2 Diagrama de frecuencias relativas (porcentajes) apiladas de las fuentes de *E. coli* por localización

La Figura 2.5.2 hace muy fácil ver que los animales de granja representan la mayor contribución de *E. coli* en Chorro Creek, mientras que los humanos son los principales responsables de la contaminación en Baywood. La distribución de las secciones en las tres barras es bastante diferente, lo que sugiere que la distribución de las fuentes de *E. coli* no es la misma en las tres localizaciones. En el Capítulo 10 aprenderemos cómo determinar si esas diferencias aparentes son lo suficientemente grandes como para ser evidencia convincente de la existencia de diferencias reales en la distribución de las fuentes de *E. coli* por localización, o si son probablemente debidas a variaciones por azar.

Relaciones numérica-categórica

En la Sección 2.4 aprendimos que los diagramas de caja son gráficos basados en solo cinco números: el mínimo, el primer cuartil, la mediana, el tercer cuartil y el máximo. Son gráficos interesantes porque son muy simples y despejados, y aun así contienen información fácil de interpretar sobre el centro, la dispersión, el sesgo e incluso los *outliers* de un conjunto de datos. Presentando **diagramas de caja uno al lado del otro** en la misma gráfica, podemos comparar datos numéricos entre varios grupos. Consideraremos a continuación una extensión del problema del crecimiento de brotes de rábanos del Ejemplo 2.4.3.

Ejemplo 2.5.3 Crecimiento de rábanos

¿Altera la exposición a la luz el crecimiento inicial de los brotes de rábano? El experimento de crecimiento de rábanos completo del Ejemplo 2.4.3 utilizó realmente un total de 42 semillas de rábano divididas aleatoriamente para recibir una de tres condiciones de iluminación en su germinación (14 semillas para cada condición de iluminación): luz las 24 horas, luz diurna (12 horas de luz y 12 horas de oscuridad en cada día) y 24 horas de oscuridad. Al final de los tres días, se midió la longitud de los brotes (mm). Por tanto, de cada brote se midieron en este estudio

dos variables: la variable categórica condición de iluminación (luz, diurna, oscuridad) y la variable numérica longitud del brote (mm). La Figura 2.5.3 muestra diagramas de caja uno al lado del otro de los datos. Los diagramas de caja hacen muy fácil comparar el crecimiento bajo las tres condiciones. Parece que la luz inhibe el crecimiento de los brotes. ¿Son las diferencias observadas en el crecimiento con respecto a las condiciones de iluminación debidas solo a variaciones por azar, o la luz realmente altera el crecimiento? Aprenderemos cómo medir numéricamente la fuerza de esta evidencia y responder a esta pregunta en los Capítulos 7 y 11. ■

Para conjuntos de datos pequeños, podemos considerar también diagramas de puntos uno al lado del otro. La Figura 2.5.4 muestra un diagrama de puntos dispersos uno al lado del otro de los datos de crecimiento de rábanos del Ejemplo 2.5.3. La «dispersión» es una opción de *software* común que añade una dispersión horizontal al diagrama, ayudando así a reducir el solapamiento de los puntos, la elección entre diagramas de caja y diagramas de puntos uno al lado del otro es cuestión de preferencia personal. Una buena opción práctica es escoger el diagrama que refleje de forma exacta los patrones en los datos de la manera más limpia posible (menos tinta sobre el papel). Para el ejemplo de crecimiento de rábanos, el diagrama de caja permite una comparación muy clara del crecimiento bajo los tres tratamientos de luz sin ocultar ninguna información revelada por el diagrama de puntos.

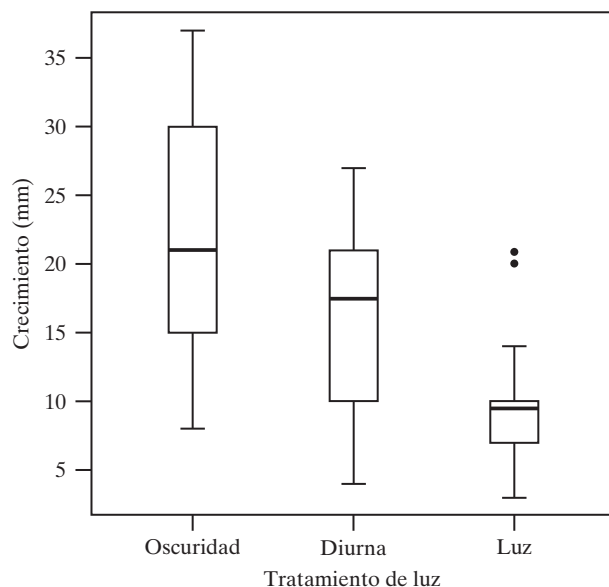


Figura 2.5.3 Diagrama de caja uno al lado del otro del crecimiento de rábanos bajo tres condiciones: oscuridad constante, luz y oscuridad a partes iguales, y luz constante

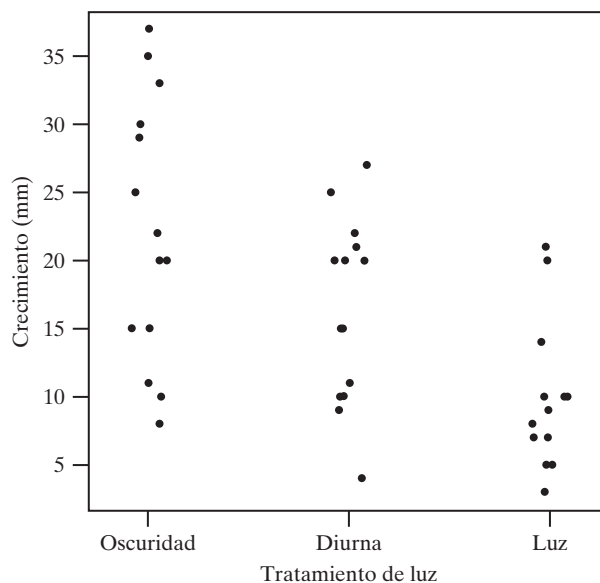


Figura 2.5.4 Diagrama de puntos dispersos uno al lado del otro del crecimiento de rábanos bajo tres condiciones: oscuridad constante, luz y oscuridad a partes iguales, y luz constante

Relaciones numérica-numérica

En cada uno de los ejemplos anteriores se consideró la comparación de la distribución de una variable (categórica o numérica) entre varios grupos (por ejemplo entre los niveles de una variable categórica). En el siguiente ejemplo ilustraremos el **diagrama de dispersión** como una herramienta para examinar la relación entre dos variables numéricas, X e Y . En un diagrama de dispersión se dibuja cada pareja (x, y) observada como un punto en el plano x - y .

Ejemplo 2.5.4 Selenio en ballenas

¿Se puede utilizar la concentración de metal en los dientes de mamíferos marinos como un bioindicador de la carga corporal? El selenio (Se) es un elemento esencial que se ha demostrado que juega un papel importante en

la protección de los mamíferos marinos contra los efectos tóxicos del mercurio (Hg) y otros metales. Se recogieron 20 ballenas beluga (*Delphinapterus leucas*) en el Mackenzie Delta, Northwest Territories, como parte de una caza tradicional anual de los Inuit³⁰. De cada una de las ballenas se tomaron dos medidas numéricas: selenio en los dientes ($\mu\text{g/g}$), y selenio en el hígado (ng/g). Las concentraciones de selenio de las ballenas se muestran en la Tabla 2.5.3. La concentración de selenio en los dientes (Y) se muestra en forma gráfica frente a la concentración de selenio en el hígado (X) en el diagrama de dispersión de la Figura 2.5.5. ■

Tabla 2.5.3 Concentración de selenio en el hígado y en los dientes de 20 belugas					
Ballena	Se en el hígado ($\mu\text{g/g}$)	Se en los dientes (ng/g)	Ballena	Se en el hígado ($\mu\text{g/g}$)	Se en los dientes (ng/g)
1	6,23	140,16	11	15,28	112,63
2	6,79	133,32	12	18,68	245,07
3	7,92	135,34	13	22,08	140,48
4	8,02	127,82	14	27,55	177,93
5	9,34	108,67	15	32,83	160,73
6	10,00	146,22	16	36,04	227,60
7	10,57	131,18	17	37,74	177,69
8	11,04	145,51	18	40,00	174,23
9	12,36	163,24	19	41,23	206,30
10	14,53	136,55	20	45,47	141,31

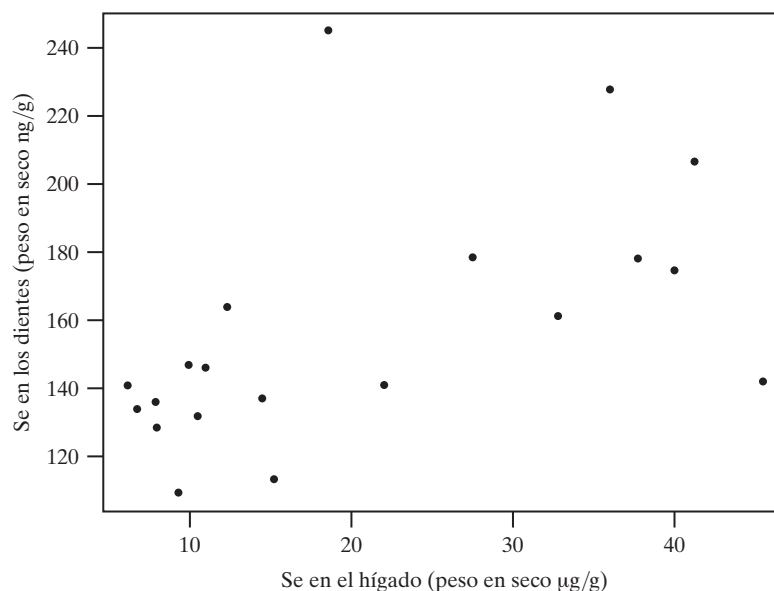


Figura 2.5.5 Diagrama de dispersión de la concentración de selenio en los dientes frente a la concentración de selenio en el hígado de 20 belugas

Los diagramas de dispersión son útiles para revelar relaciones entre variables numéricas. En la Figura 2.5.6 se han añadido dos líneas al diagrama de dispersión de la concentración de selenio en ballenas de la Figura 2.5.5. Esas dos líneas son para resaltar la tendencia creciente de los datos: la concentración de selenio en los dientes tiende a incrementarse con la concentración de selenio en el hígado. La línea discontinua se denomina **suavizado lowess** (*locally weighted scatterplot smoothing* – suavizado del diagrama de dispersión mediante promediado local), y la línea continua se denomina **recta de regresión**. Muchos paquetes de *software* permiten añadir fácilmente esas líneas a un diagrama de dispersión. El suavizado *lowess* es particularmente útil para visualizar relaciones curvas o no lineales en los datos, mientras que la recta de regresión se utiliza para resaltar una tendencia lineal. Hablando en términos generales, deberíamos escoger solo una de ellas para mostrar nuestra gráfica. En este caso, como el patrón es bastante lineal (el suavizado *lowess* es bastante recto), escogeríamos la recta de regresión continua. En el Capítulo 12 aprenderemos a identificar la ecuación de la recta de regresión que mejor resume el comportamiento de los datos y a determinar si la tendencia aparente de los datos es probable que sea debida solo al azar o si existe evidencia de una relación real entre X e Y .

Además de revelar relaciones entre dos variables numéricas, los diagramas de dispersión ayudan también a revelar *outliers* que podrían pasar de otra forma inadvertidos en diagramas univariados (por ejemplo, en histogramas, diagramas de caja simples, etc.). El punto coloreado de la Figura 2.5.6 está lejos de la dispersión de los otros puntos. El valor de X de este punto no es inusual en absoluto, e incluso el valor de Y , aunque grande, no parece extremo. Sin embargo, el diagrama de dispersión muestra que la pareja particular (x, y) de esta ballena es inusual.

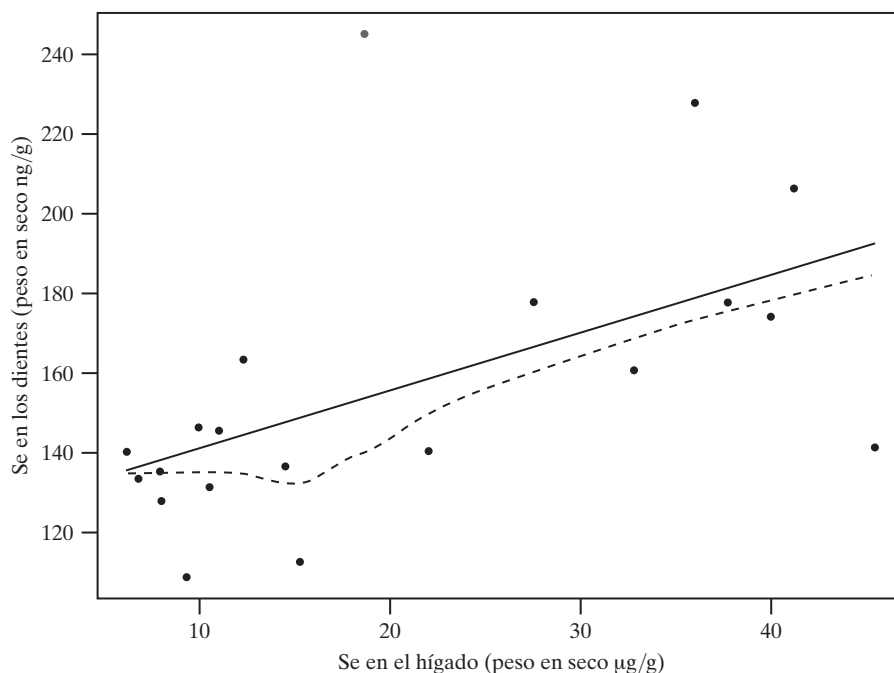


Figura 2.5.6 Diagrama de dispersión de la concentración de selenio en los dientes frente a la concentración de selenio en el hígado de 20 ballenas con la recta de regresión (continua) y la línea de suavizado *lowess* (discontinua) y un *outlier* marcado en gris

Ejercicios 2.5.1-2.5.3

2.5.1 Las dos pinzas del bogavante (*Homarus americanus*) son idénticas en las etapas juveniles. Sin embargo, al llegar a la etapa adulta, las dos pinzas normalmente se han diferenciado en una pinza resistente denominada «trituratora» y una pinza más delgada denominada «cortadora». En un estudio sobre el proceso de diferenciación, 26 animales jóvenes se criaron en bandejas de plástico suave y 18 se criaron en bandejas que contenían trocitos de concha de ostra (que podían utilizar para ejercitar sus pinzas). Otros 23 animales se criaron en bandejas que contenían solo un trocito de concha de ostra. Las configuraciones de las pinzas de todos los animales cuando llegaron al estado adulto se resumen en la tabla³¹.

Tratamiento	Configuración de las pinzas		
	Trituratora derecha, cortadora izquierda	Cortadora derecha, trituratora izquierda	Cortadoras derecha e izquierda (sin trituratora)
Trocitos de concha de ostra	8	9	1
Plástico suave	2	4	20
Un trocito de concha de ostra	7	9	7

- (a) Cree un diagrama de frecuencias de barras apiladas para mostrar estos datos.
 (b) Cree un diagrama de frecuencias relativas de barras apiladas para mostrar estos datos.
 (c) De los dos diagramas que ha creado en los apartados (a) y (b), ¿cuál es más útil para comparar las configuraciones de las pinzas con los tres tratamientos? ¿Por qué?

2.5.2 ¿Difiere con la altitud en California la longitud (mm) de la ardilla de tierra de manto dorado (*Spermophilus lateralis*)? Un estudiante graduado capturó ardillas en cuatro localizaciones de California. De sur a norte, las localizaciones eran Hemet, Big Bear, Susanville y Loop Hill³².

Hemet	Big Bear	Susanville	Loop Hill
263	274	345	273
256	256	272	291
251	249	263	278
242	264	260	281
248		271	
281			

- (a) Cree diagramas de puntos uno al lado del otro de los datos. Considere la geografía de esas cuatro localizaciones al realizar su gráfica. ¿Es lo más apropiado situar las localizaciones en orden alfabético o hay una manera mejor de ordenar las categorías de localizaciones?
 (b) Cree diagramas de caja uno al lado del otro de los datos. Considere de nuevo la geografía de esas cuatro localizaciones al realizar su gráfica.
 (c) De los dos diagramas creados en los apartados (a) y (b), ¿cuál preferiría y por qué?

2.5.3 El serbal (*Sorbus aucuparia*) es un árbol que crece en un amplio intervalo de altitudes. Para estudiar cómo se adapta el árbol a los diferentes hábitats, unos investigadores recogieron ramitas que tuvieran brotes de 12 árboles que crecían en diversas altitudes en North Angus, Escocia. Los brotes se llevaron al laboratorio y se realizaron medidas sobre la tasa de respiración en oscuridad. La tabla que sigue muestra la altitud del origen (en metros) de cada lote de brotes y la tasa de respiración en oscuridad (expresada como μl de oxígeno por hora por mg de peso de tejido en seco)³³.

Árbol	Altitud de origen (m)	Tasa de respiración ($\mu\text{l/hr} \cdot \text{mg}$)
1	90	0,11
2	230	0,20
3	240	0,13
4	260	0,15
5	330	0,18
6	400	0,16
7	410	0,23
8	550	0,18
9	590	0,23
10	610	0,26
11	700	0,32
12	790	0,37

- (a) Cree un diagrama de dispersión de los datos.
 (b) Si su *software* lo permite, añada una recta de regresión para resumir la tendencia.
 (c) Si su *software* lo permite, cree un diagrama de dispersión con un suavizado *lowess* para resumir la tendencia.

2.6 Medidas de dispersión

Hemos considerado las formas y los centros de las distribuciones, pero una buena descripción de la distribución debería también caracterizar el grado de dispersión de la misma (¿son las observaciones de la muestra casi todas iguales, o difieren sustancialmente?). En la Sección 2.4 definimos el intervalo intercuartílico, que es una medida de dispersión. Consideraremos ahora otras medidas de dispersión: el rango, la desviación típica y el coeficiente de variación.

El rango

El **rango** muestral es la diferencia entre las observaciones máxima y mínima de la muestra. Presentamos a continuación un ejemplo.

Ejemplo 2.6.1 Presión sanguínea

Las presiones sanguíneas sistólicas (mm Hg) de seis hombres de mediana edad que se presentaron en el Ejemplo 2.4.1 fueron las siguientes:

113 124 124 132 146 151 170

Para estos datos, el rango muestral es

$$170 - 113 = 57 \text{ mm Hg}$$

El rango es fácil de calcular, pero es muy sensible a los valores extremos. Es decir, no es robusto. Si el máximo de la muestra de la presión sanguínea hubiera sido 190 en vez de 170, el rango habría cambiado de 57 a 77.

En la Sección 2.4 definimos el intervalo intercuartílico (IQR) como la diferencia entre los cuartiles. A diferencia del rango, el IQR es robusto. El IQR de la presión sanguínea es $151 - 124 = 27$. Si el máximo de la muestra de la presión sanguínea hubiera sido 190 en vez de 170, el IQR no habría cambiado; seguiría siendo 27.

La desviación típica

La desviación típica es la medida de dispersión clásica y más ampliamente utilizada. Recuerdese que una *desviación* es la diferencia entre una observación y la media muestral:

$$\text{desviación} = \text{observación} - \bar{y}$$

La desviación típica de la muestra, o **desviación típica** muestral, se determina combinando las desviaciones de una forma especial, como se indica en el siguiente recuadro.

La desviación típica muestral La desviación típica muestral se denota como s y se define mediante la siguiente fórmula:

$$s = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}}$$

En esta fórmula, la expresión $\sum_{i=1}^n (y_i - \bar{y})^2$ indica la suma de las desviaciones al cuadrado.

Por tanto, para obtener la desviación típica de una muestra, primero se obtienen las desviaciones. Seguidamente

1. Se elevan al cuadrado.
2. Se suman.
3. Se dividen por $n - 1$.
4. Se toma la raíz cuadrada.

Para ilustrar el uso de la fórmula, hemos escogido un conjunto de datos que es especialmente simple de manejar porque su media es un entero.

Ejemplo 2.6.2 Crecimiento de crisantemos

En un experimento sobre crisantemos, un botánico midió el alargamiento del tallo (mm en 7 días) de cinco plantas que crecían en el mismo banco del invernadero. Los resultados fueron los siguientes³⁴:

$$76 \quad 72 \quad 65 \quad 70 \quad 82$$

Los datos están tabulados en la primera columna de la Tabla 2.6.1. La media muestral es

$$\bar{y} = \frac{365}{5} = 73 \text{ mm}$$

Las desviaciones $(y_i - \bar{y})$ están tabuladas en la segunda columna de la Tabla 2.6.1. La primera observación está 3 mm por encima de la media, la segunda está 1 mm por debajo de la media, y así sucesivamente.

La tercera columna de la Tabla 2.6.1 muestra que la suma de las desviaciones al cuadrado es

$$= \sum_{i=1}^n (y_i - \bar{y})^2 = 164$$

Tabla 2.6.1 Ilustración de la fórmula de la desviación típica muestral		
Observación (y_i)	Desviación ($y_i - \bar{y}$)	Desviación al cuadrado ($(y_i - \bar{y})^2$)
76	3	9
72	-1	1
65	-8	64
70	-3	9
82	9	81
Suma $365 = \sum_{i=1}^n y_i$	0	$164 = \sum_{i=1}^n (y_i - \bar{y})^2$

Como $n = 5$, las desviación típica es

$$s = \sqrt{\frac{164}{4}} = \sqrt{41} = 6,4 \text{ mm}$$

Nótese que las unidades de s (mm) son las mismas que las unidades de Y . Esto es porque hemos elevado al cuadrado las desviaciones y después hemos tomado la raíz cuadrada. ■

La **varianza** muestral, denotada como s^2 , es simplemente la desviación típica al cuadrado: $\text{varianza} = s^2$. Por tanto, $s = \sqrt{\text{varianza}}$.

Ejemplo 2.6.3 Crecimiento de crisantemos

La varianza de los datos de crecimiento de crisantemos es

$$s^2 = 41 \text{ mm}^2$$

Nótese que las unidades de la varianza (mm^2) no son las mismas que las unidades de Y . ■

Una abreviatura Frecuentemente abreviaremos «desviación típica» como «DT»; en las fórmulas se utilizará el símbolo « s ».

Interpretación de la definición de s

El módulo (es decir, no considerando el signo) de cada desviación ($y_i - \bar{y}$) se puede interpretar como la *distancia* de la correspondiente observación a la media muestral $\bar{y}$. La Figura 2.6.1 muestra una gráfica de los datos de crecimiento de crisantemos (Ejemplo 2.6.2) donde se ha marcado cada distancia.

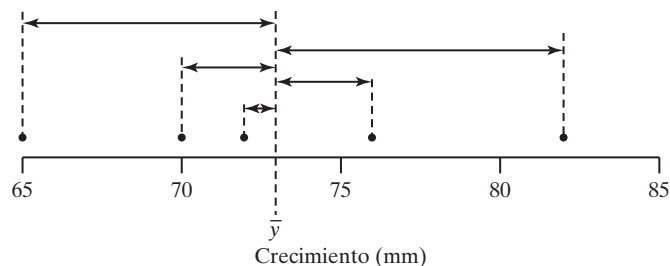


Figura 2.6.1 Gráfica de los datos de crecimiento de crisantemos con las desviaciones indicadas en forma de distancias

A partir de la fórmula de s , puede verse que cada desviación contribuye a la DT. Por tanto, una muestra del mismo tamaño pero con menor dispersión tendrá una DT menor, como se ilustra en el siguiente ejemplo.

Ejemplo 2.6.4 Crecimiento de crisantemos

Si los datos de crecimiento de crisantemos del Ejemplo 2.6.2 hubieran cambiado a

75 72 73 75 70

entonces la media es la misma ($\bar{y} = 73$ mm), pero la DT es menor ($s = 2,1$ mm), debido a que las observaciones están más cerca de la media. La dispersión relativa de las dos muestras se puede ver fácilmente en la Figura 2.6.2.

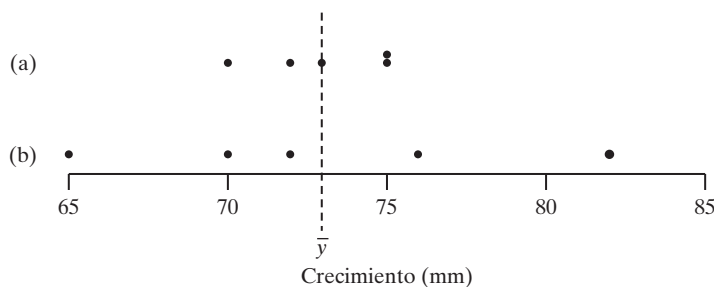


Figura 2.6.2 Dos muestras de datos de crecimiento de crisantemos con la misma media, pero con diferentes desviaciones típicas (a) $s = 2,1$ mm; (b) $s = 6,3$ mm.

Observemos con más detalle la forma en que se combinan las desviaciones para formar la DT. La fórmula indica que hay que dividir por $(n - 1)$. Si el divisor fuera n en vez de $(n - 1)$, entonces la cantidad dentro de la raíz cuadrada sería el promedio (la media) de las desviaciones al cuadrado. A menos que n sea muy pequeño, el incremento debido a dividir por $(n - 1)$ en vez de por n no es muy grande, de forma que la DT se puede interpretar aproximadamente como

$$s \approx \sqrt{\text{valor medio muestral de } (y_i - \bar{y})^2}$$

Por tanto, es aproximadamente apropiado ver la DT como la distancia «típica» de las observaciones a su media.

¿Por qué $n - 1$? Como dividir por n parece más natural, podemos preguntarnos por qué la fórmula de la DT divide por $(n - 1)$. Nótese que la suma de las desviaciones es siempre cero. Por tanto, una vez que se han calculado las primeras $n - 1$ desviaciones, la última desviación está restringida. Esto significa que una muestra con n observaciones hay solo $n - 1$ unidades de información con respecto a la desviación del promedio. La cantidad $n - 1$ se denomina los **grados de libertad** de la desviación típica o de la varianza. Podemos dar también una justificación intuitiva de por qué se usa $n - 1$ considerando el caso extremo de $n = 1$, como en el ejemplo siguiente.

Ejemplo 2.6.5 Crecimiento de crisantemos

Supongamos que el experimento sobre crecimiento de crisantemos del Ejemplo 2.6.2 hubiera incluido solo una planta, de forma que la muestra consistiera únicamente en la observación

$$73$$

Para esta muestra, $n = 1$ e $\bar{y} = 73$. Sin embargo, la fórmula de la DT falla (ya que resulta $\frac{0}{0}$), por lo que no se puede calcular la DT. Esto es razonable, ya que la muestra no da ninguna información sobre la variabilidad del crecimiento de los crisantemos bajo las condiciones experimentales. Si en la fórmula de DT se hubiera dividido por n , habríamos obtenido una DT de cero, sugiriendo que hay poca o ninguna variabilidad. Esa conclusión parece poco justificable por la observación de una única planta. ■

El coeficiente de variación

El **coeficiente de variación** en la desviación típica expresada como un porcentaje de la media: coeficiente de variación = $\frac{s}{\bar{y}} \times 100 \%$. Presentamos a continuación un ejemplo.

Ejemplo 2.6.6 Crecimiento de crisantemos

En el Ejemplo 2.6.2 de crecimiento de crisantemos, obtuvimos $\bar{y} = 73,0$ mm y $s = 6,4$ mm. Por tanto,

$$\frac{s}{\bar{y}} \times 100 \% = \frac{6,4}{73,0} \times 100 \% = 0,088 \times 100 \% = 8,8 \%$$

El coeficiente de variación de la muestra es 8,8 %. Por tanto, la desviación típica es el 8,8 % de la media. ■

Nótese que el coeficiente de variación no es afectado por cambios multiplicativos de escala. Por ejemplo, si los datos de los crisantemos se expresaran en pulgadas en vez de milímetros, entonces tanto $\bar{y}$ como s se expresarían en pulgadas, y el coeficiente de variación no cambiaría. Debido a su inmunidad a los cambios de escala, el coeficiente de variación es una medida útil para comparar las dispersiones de dos o más variables que se miden en escalas diferentes.

Ejemplo 2.6.7 Alturas y pesos de niñas

Como parte de la Berkeley Guidance Study³⁵, se midieron las alturas (en cm) y los pesos (en kg) de 13 niñas de dos años. A la edad de dos años, la altura media fue de 86,6 cm y la DT fue de 2,9 cm. Por tanto, el coeficiente de variación de la altura a la edad de dos años es

$$\frac{s}{\bar{y}} \times 100 \% = \frac{2,9}{86,6} \times 100 \% = 0,033 \times 100 \% = 3,3 \%$$

El promedio del peso a la edad de dos años fue de 12,6 kg y la DT fue de 1,4 kg. Por tanto, el coeficiente de variación del peso a la edad de dos años es

$$\frac{s}{\bar{y}} \times 100 \% = \frac{1,4}{12,6} \times 100 \% = 0,111 \times 100 \% = 11,1 \%$$

Hay considerablemente más variabilidad en el peso que en la altura, cuando expresamos cada medida de variabilidad como un porcentaje de la media. La DT del peso es un porcentaje bastante grande del peso medio, pero la DT de la altura es un porcentaje más bien pequeño de la altura media. ■

Representación de medidas de dispersión

El rango y el intervalo intercuartílico son fáciles de interpretar. El rango es la dispersión de todas las observaciones y el intervalo intercuartílico es la dispersión de (aproximadamente) el 50 % central de las observaciones. En términos del histograma de un conjunto de datos, el rango se puede ver como (aproximadamente) la anchura del histograma. Los cuartiles son (aproximadamente) los valores que dividen el área en cuatro partes iguales y el intervalo intercuartílico es la distancia entre los cuartiles primero y tercero. El ejemplo siguiente ilustra estas ideas.

Ejemplo 2.6.8 Ganancia diaria de ganado

Se evaluó el rendimiento del ganado vacuno midiendo su ganancia de peso durante un periodo de prueba de 140 días con una dieta estándar. La Tabla 2.6.2 muestra las ganancias diarias promedio (kg/día) de 39 toros de la misma raza (Charolais). Las observaciones se presentan en orden creciente³⁶. Los valores van desde 1,18 kg/día hasta 1,92 kg/día. La Figura 2.6.3 muestra un histograma de los datos, el rango, los cuartiles y el intervalo intercuartílico (IQR). El área sombreada representa (aproximadamente) el 50 % central de las observaciones. ■

1,18	1,24	1,29	1,37	1,41	1,51	1,58	1,72
1,20	1,26	1,33	1,37	1,41	1,53	1,59	1,76
1,23	1,27	1,34	1,38	1,44	1,55	1,64	1,83
1,23	1,29	1,36	1,40	1,48	1,57	1,64	1,92
1,23	1,29	1,36	1,41	1,50	1,58	1,65	

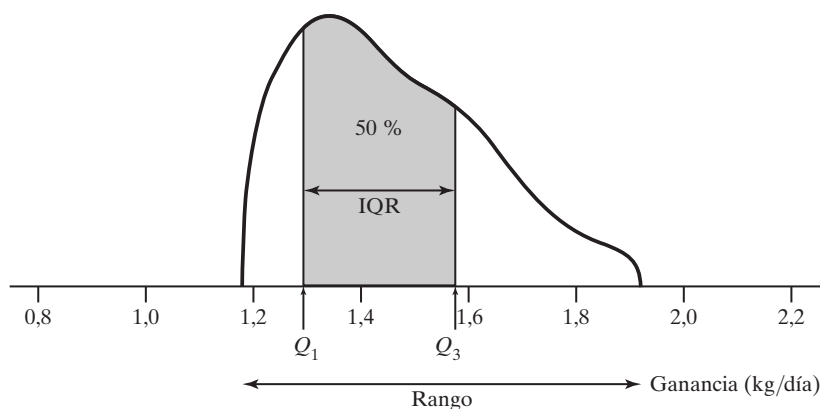


Figura 2.6.3 Histograma suavizado de 39 medidas de ganancia diarias, indicando el rango, los cuartiles y el intervalo intercuartílico (IQR). El área sombreada representa aproximadamente el 50 % de las observaciones

Representación de la desviación típica

Hemos visto que la DT es una medida combinada de las distancias de las observaciones a su media. Es natural preguntarse cuántas observaciones están dentro de un intervalo de ± 1 DT alrededor de la media, cuántas están dentro de un intervalo de ± 2 DT alrededor de la media, y así sucesivamente. El siguiente ejemplo explora esta cuestión.

Ejemplo 2.6.9 Ganancia diaria de ganado

Para los datos de ganancia diaria del Ejemplo 2.6.8, la media es $\bar{y} = 1,445$ kg/día y la DT es $s = 0,183$ kg/día. En la Figura 2.6.4 se han marcado sobre un histograma de los datos los intervalos $\bar{y} \pm s$, $\bar{y} \pm 2s$ e $\bar{y} \pm 3s$. El intervalo $\bar{y} \pm s$ es

$$1,445 \pm 0,183 \text{ o } 1,262 \text{ hasta } 1,628$$

Observando la Tabla 2.6.2 se puede verificar que este intervalo contiene 25 de las 39 observaciones. Por tanto, $\frac{25}{39}$ o el 64 % de las observaciones están dentro de un intervalo de ± 1 DT alrededor de la media. El área correspondiente está sombreada en la Figura 2.6.4. El intervalo $\bar{y} \pm 2s$ es

$$1,445 \pm 0,366 \text{ o } 1,079 \text{ hasta } 1,811$$

Este intervalo contiene $\frac{37}{39}$ o el 95 % de las observaciones. Se puede verificar que el intervalo $\bar{y} \pm 3s$ contiene todas las observaciones. ■

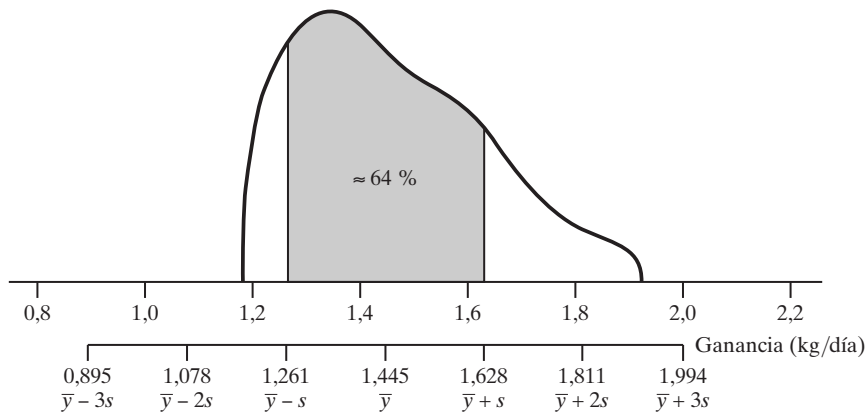


Figura 2.6.4 Histograma de los datos de ganancia diaria que muestra los intervalos de 1, 2 y 3 desviaciones típicas alrededor de la media. El área sombreada representa aproximadamente el 64 % de las observaciones

Los porcentajes del Ejemplo 2.6.9 resultan ser bastante típicos de distribuciones observadas en ciencias.

Porcentajes típicos: la regla empírica

Para distribuciones con «forma buena», es decir, distribuciones unimodales que no estén demasiado sesgadas y cuyas colas no sean demasiado largas y cortas, en general podemos esperar encontrar:

Aproximadamente 68 % de las observaciones dentro de una distancia de ± 1 DT de la media.

Aproximadamente el 95 % de las observaciones dentro de una distancia de ± 2 DT de la media.

>99 % de las observaciones dentro de una distancia de ± 3 DT de la media.

Los porcentajes típicos nos permiten construir una imagen mental aproximada de una distribución de frecuencias sabiendo solo la media y la DT. (El valor de 68 % parece haber aparecido de la nada. Su origen quedará claro en el Capítulo 4).

Estimación de la DT a partir de un histograma

La regla empírica nos proporciona una forma de construir una imagen mental aproximada de una distribución de frecuencias conociendo solo la media y la DT. Podemos visualizar un histograma centrado en la media y extendiéndose un poco más de 2 DT en cada dirección. Por supuesto, la distribución real podría no ser simétrica, pero nuestra imagen mental aproximada a menudo será bastante exacta.

Pensando sobre esto a la inversa, podemos ver un histograma y estimar la DT. Para ello, necesitamos estimar los extremos del intervalo que esté centrado en la media y que contenga aproximadamente el 95 % de los datos. La regla empírica implica que este intervalo es aproximadamente el mismo que $(\bar{y} - 2s, \bar{y} + 2s)$, por lo que la longitud del intervalo debería ser aproximadamente cuatro veces la DT:

$$(\bar{y} - 2s, \bar{y} + 2s) \text{ tiene una longitud de } 2s + 2s = 4s$$

Esto significa que

$$\text{longitud del intervalo} = 4s$$

Por lo que

$$\text{estimación de } s = \frac{\text{longitud del intervalo}}{4}$$

Por supuesto, nuestra estimación visual del intervalo que abarca el 95 % de los datos alrededor de la mitad de la distribución podría ser errónea. Es más, la regla empírica trabaja mejor con distribuciones que son simétricas. Por tanto, este método de estimar la DT producirá únicamente una estimación general. El método funciona mejor cuando la distribución es bastante simétrica, pero funciona razonablemente bien incluso si la distribución es algo sesgada.

Ejemplo 2.6.10 Pulsaciones después de hacer ejercicio

Un grupo de 28 adultos realizó un ejercicio moderado durante cinco minutos y después se midieron sus pulsaciones. La Figura 2.6.5 muestra la distribución de los datos³⁷. Podemos ver que aproximadamente el 95 % de las observaciones están entre 75 y 125*. Por tanto, un intervalo de longitud 50 ($50 = 125 - 75$) comprende el 95 % de los datos situados en la mitad. Teniendo esto en cuenta, podemos estimar que la DT es $\frac{50}{4} = 12,5$. La DT real es 13,4, que no está muy lejos de nuestra estimación. ■

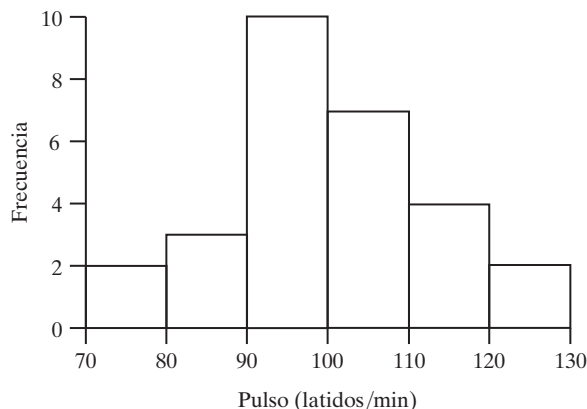


Figura 2.6.5 Pulsaciones después de un ejercicio moderado en un grupo de adultos

Los porcentajes típicos dados por la regla empírica pueden tener bastante error si la muestra es pequeña o si la forma de la distribución de frecuencias no es «buena». Por ejemplo, los datos de tiempos de canto de grillos (Tabla 2.3.1 y Figura 2.3.4) tienen $s = 4,4$ mm y el intervalo $\bar{y} \pm s$ contiene el 90 % de las observaciones. Esto es un valor mucho más alto que el «típico» del 68 %, ya que la DT ha aumentado por la cola larga e irregular de la distribución.

Comparación de medidas de dispersión

La dispersión de los datos en una muestra se puede describir mediante la desviación típica, el rango o el intervalo intercuartílico. El rango es fácil de entender, pero puede ser una medida descriptiva pobre debido a que depende solo de

* Es difícil ver de forma exacta utilizando un histograma dónde está el 95 % de los datos situados en la mitad, pero se trata no solo de una estimación visual, sin el objetivo de producir un valor exacto. Nuestras estimaciones visuales de la DT podrían diferir entre sí, pero todas tendrían valores relativamente cercanos.

las colas extremas de la distribución. Por el contrario, el intervalo intercuartílico describe la dispersión del «cuerpo» central de la distribución. La desviación típica tiene en cuenta todas las observaciones y se puede interpretar aproximadamente en términos de la dispersión de las observaciones alrededor de su media. Sin embargo, la DT puede estar sobrevalorada por observaciones en los extremos de las colas. El intervalo intercuartílico es una medida resistente, mientras que la DT no es resistente. Por supuesto, el rango es extremadamente no resistente.

La interpretación descriptiva de la DT es menos directa que la del rango o el intervalo intercuartílico. No obstante, la DT es la base de la mayoría de los métodos estadísticos clásicos estándar. La DT goza de este *status* por varias razones técnicas, entre las que se encuentra su eficiencia en ciertas situaciones.

Los desarrollos de los capítulos posteriores harán énfasis en métodos estadísticos clásicos en los que la media y la DT juegan un papel central. En consecuencia, en este libro nos basaremos primordialmente en la media y la DT más que en otras medidas descriptivas.

Ejercicios 2.6.1-2.6.16

2.6.1 Calcule la desviación típica de cada una de estas muestras ficticias:

- (a) 16, 13, 18, 13.
- (b) 38, 30, 34, 38, 35.
- (c) 1, -1, 5, -1.
- (d) 4, 6, -1, 4, 2.

2.6.2 Calcule la desviación típica de cada una de estas muestras ficticias:

- (a) 8, 6, 9, 48.
- (b) 4, 7, 5, 4.
- (c) 9, 2, 6, 7, 6.

2.6.3

- (a) Invente una muestra de tamaño 5 en la que las desviaciones ($y_i - \bar{y}$) sean -3, -1, 0, 2, 2.
- (b) Calcule la desviación típica de su muestra.
- (c) ¿Debería dar todo el mundo la misma respuesta al apartado (b)? ¿Por qué?

2.6.4 Cuatro parcelas de tierra, cada una de ellas de 346 pies cuadrados, fueron sembradas con la misma variedad de trigo («Beau»). Las cosechas de las parcelas (lb) fueron las siguientes³⁸:

35,1 30,6 36,9 29,8

- (a) Calcule la media y la desviación típica.
- (b) Calcule el coeficiente de variación.

2.6.5 Un fisiólogo de plantas hizo crecer semilleros de abedul en un invernadero y midió el contenido de ATP de sus raíces. (Véase el Ejemplo 1.1.3). Los resultados (nmol de ATP/mg de tejido) se presentan a continuación para los cuatro semilleros que se habían cuidado de la misma forma³⁹:

1,45 1,19 1,05 1,07

- (a) Calcule la media y la desviación típica.
- (b) Calcule el coeficiente de variación.

2.6.6 10 pacientes con la presión sanguínea alta participaron en un estudio para evaluar la efectividad del medicamento Timolol para reducir su presión sanguínea. La tabla

que sigue muestra las medidas de presión sanguínea sistólica tomadas antes y después de dos semanas de tratamiento con Timolol⁴⁰. Calcule la media y la desviación típica del *cambio* en la presión sanguínea (nótese que algunos valores son negativos).

Paciente	Presión sanguínea		
	Antes	Después	Cambio
1	172	159	-13
2	186	157	-29
3	170	163	-7
4	205	207	2
5	174	164	-10
6	184	141	-43
7	178	182	4
8	156	171	15
9	190	177	-13
10	168	138	-30

2.6.7 La dopamina es un compuesto químico que participa en la transmisión de señales en el cerebro. Un farmacólogo midió la cantidad de dopamina en el cerebro de siete ratas. Los niveles de dopamina (nmoles/g) fueron los siguientes⁴¹:

6,8 5,3 6,0 5,9 6,8 7,4 6,2

- (a) Calcule la media y la desviación típica.
- (b) Determine la mediana y el intervalo intercuartílico.
- (c) Calcule el coeficiente de variación.
- (d) Sustituya la observación 7,4 por 10,4 y repita los apartados (a) y (b). ¿Qué medidas descriptivas muestran resistencia y cuáles no?

2.6.8 En un estudio sobre el lagarto *Sceloporus occidentalis*, unos biólogos midieron la distancia (m) recorrida en dos minutos por 15 animales. Los resultados (dispuestos en orden creciente) fueron los siguientes⁴²:

18,4 22,2 24,5 26,4 27,5 28,7 30,6 32,9
32,9 34,0 34,8 37,5 42,1 45,5 45,5

- (a) Determine los cuartiles y el intervalo intercuartílico.
(b) Determine el rango.

2.6.9 Haciendo referencia a los datos de distancia recorrida del Ejercicio 2.6.8, la media muestral es 32,23 m y la DT es 8,07 m. ¿Qué porcentaje de observaciones están dentro de

- (a) 1 DT alrededor de la media?
(b) 2 DT alrededor de la media?

2.6.10 Compare los resultados del Ejercicio 2.6.9 con las predicciones de la regla empírica.

2.6.11 Se presentan en orden creciente los niveles en suero (U/l) de creatina fosfoquinasa (CFK) de 36 hombres saludables (son los datos del Ejemplo 2.2.6):

25	62	82	95	110	139
42	64	83	95	113	145
48	67	84	100	118	151
57	68	92	101	119	163
58	70	93	104	121	201
60	78	94	110	123	203

La media muestral del nivel de CFK es 98,3 U/l y la DT es 40,4 U/l. ¿Qué porcentaje de las observaciones están dentro de

- (a) 1 DT alrededor de la media?
(b) 2 DT alrededor de la media?
(c) 3 DT alrededor de la media?

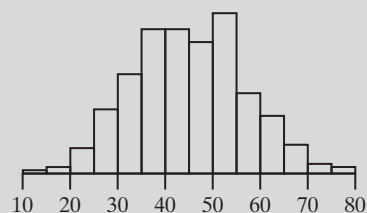
2.6.12 Compare los resultados del Ejercicio 2.6.11 con las predicciones de la regla empírica.

2.6.13 Las niñas del Berkeley Guidance Study (Ejemplo 2.6.7), que fueron medidas a la edad de dos años, fueron medidas de nuevo a la edad de nueve años. Por supuesto la altura y el peso medios eran mucho mayores a la edad de nueve años que a la edad de dos. Asimismo, las DT de la altura y del

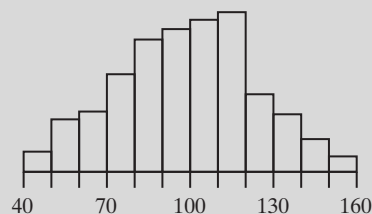
peso eran mucho mayores a la edad de nueve años que a la edad de dos. ¿Pero qué ocurrió con el coeficiente de variación de la altura y el coeficiente de variación del peso? Sucedió que uno de ellos aumentó una cantidad moderada de la edad de dos a la edad de nueve, pero en la otra variable el incremento del coeficiente de variación fue bastante grande. ¿De qué variable, peso o altura, se podría esperar que el coeficiente de variación cambiará más entre la edad de dos y la edad de nueve? ¿Por qué? (*Sugerencia:* piense en cómo los factores genéticos influyen en la altura y el peso y en cómo los factores ambientales influyen en la altura y el peso).

2.6.14 Considere las 13 niñas mencionadas en el Ejemplo 2.6.7. A la edad de 18 su altura media era de 166,3 cm y la DT de sus alturas era de 6,8 cm. Calcule el coeficiente de variación.

2.6.15 Se presenta a continuación un histograma. Estime la media y la DT de la distribución.



2.6.16 Se presenta a continuación un histograma. Estime la media y la DT de la distribución.



2.7 Efectos de transformación de variables (opcional)

Algunas veces, cuando estamos trabajando con un conjunto de datos, podemos encontrar conveniente transformar una variable. Por ejemplo, podríamos convertir de pulgadas a centímetros o de grados Fahrenheit a grados centígrados. La transformación o reexpresión de una variable Y significa sustituir Y por una nueva variable, por ejemplo Y' . Para trabajar más cómodamente con los datos, es útil conocer cómo son afectadas las características de una distribución si se transforma la variable observada.

Las transformaciones más simples son las transformaciones **lineales**, que se denominan así porque la gráfica de Y con respecto a Y' es una línea recta. Una razón familiar para utilizar una transformación lineal es un cambio en la escala de medida, como se ilustra en los dos ejemplos siguientes.

Ejemplo 2.7.1 Peso

Supongamos que Y representa el peso de un animal en kg, y decidimos expresar el peso en lb. Entonces

$$Y = \text{peso en kg}$$

$$Y' = \text{peso en lb}$$

por tanto,

$$Y' = 2,2Y$$

Esto es una transformación **multiplicativa**, ya que Y' se calcula a partir de Y multiplicándola por el valor constante en 2,2. ■

Ejemplo 2.7.2 Temperatura corporal

Se realizaron medidas de la temperatura del cuerpo (temperatura caminando) en 47 mujeres⁴³. Algunas observaciones típicas Y , en °C, fueron

$$Y: 36,23, 36,41, 36,77, 36,15, \dots$$

Supongamos que convertimos esos datos de °C a °F, y denominamos Y' a la nueva variable:

$$Y': 97,21, 97,54, 98,19, 97,07, \dots$$

La relación entre Y e Y' es

$$Y' = 1,8Y + 32$$

La combinación de cambios **aditiva** (+32) y multiplicativa ($\times 1,8$) indica una relación lineal. ■

Otra razón para utilizar transformaciones lineales es la **codificación**, que significa transformar los datos por comodidad al manejar los números. Se presenta a continuación un ejemplo.

Ejemplo 2.7.3 Temperatura corporal

Consideremos los datos de temperatura del Ejemplo 2.7.2. Si restamos 36 a cada observación, los datos se convierten en

$$0,23, 0,41, 0,77, 0,15, \dots$$

Esto es una codificación aditiva, ya que hemos añadido un valor constante (-36) a cada observación. Supongamos que ahora transformamos los datos en la forma

$$23, 41, 77, 15, \dots$$

Este paso de la codificación es multiplicativo, ya que cada observación se multiplica por un valor constante (100). ■

Como ilustran los ejemplos anteriores, una transformación lineal consiste en (1) multiplicar todas las observaciones por una constante o (2) sumar una constante a todas las observaciones o (3) ambas cosas.

Como afectan las transformaciones lineales a la distribución de frecuencias

Una transformación lineal de los datos no cambia la forma esencial de su distribución de frecuencias. Modificando adecuadamente la escala del eje horizontal, puede hacerse que el histograma transformado sea idéntico al histograma original. El Ejemplo 2.7.4 ilustra esta idea.

Ejemplo 2.7.4 Temperatura corporal

La Figura 2.7.1 muestra la distribución de 47 medidas de temperatura que se han transformado restando primero 36 a cada observación y multiplicando después las observaciones por 100 (como en los Ejemplos 2.7.2 y 2.7.3). Es decir, $Y' = (Y - 36) \times 100$. La figura muestra que las dos distribuciones se pueden representar mediante el mismo histograma con diferentes escalas horizontales. ■

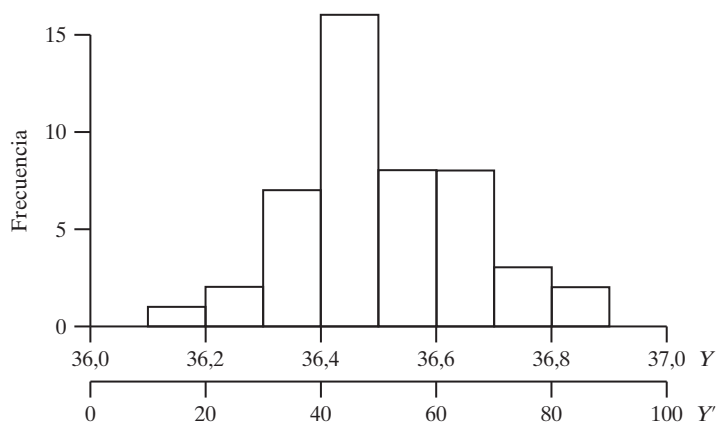


Figura 2.7.1 Distribución de 47 medidas de temperatura donde se muestran la escala original y la escala transformada linealmente

Cómo afectan las transformaciones lineales a $\bar{y}$ y a s

El efecto de una transformación lineal sobre $\bar{y}$ es «natural». Es decir, **bajo una transformación lineal**, $\bar{y}$ cambia como Y . Por ejemplo, si las temperaturas se transforman de °C a °F, entonces la media se transforma de forma similar:

$$Y' = 1,8Y + 32; \quad \text{por tanto} \quad \bar{y}' = 1,8\bar{y} + 32$$

El efecto sobre s de multiplicar Y por una constante positiva es «natural». Si $Y' = c \times Y$, con $c > 0$, entonces $s' = c \times s$. Por ejemplo, si se transforman pesos de kg a lb, la DT se convierte de forma similar: $s' = 2,2s$. Si $Y' = c \times Y$, con $c < 0$, entonces $s' = -c \times s$. En general, $Y' = c \times Y$, entonces $s' = |c| \times s$.

Sin embargo, una transformación aditiva no afecta a s . Si sumamos o restamos una constante, no cambiamos la dispersión de la distribución, y por lo tanto s no cambia. Así, por ejemplo, *no* transformaríamos la DT de los datos de temperatura de °C en °F de la misma forma en la que se transforma cada observación. La DT se multiplicaría por 1,8, pero *no* se sumaría 32. El hecho de que la DT no cambie con transformaciones aditivas parecerá menos sorprendente si se recuerda (de la definición) que s depende solo de las desviaciones $(y_i - \bar{y})$, y esas desviaciones no cambian con transformaciones aditivas. El siguiente ejemplo ilustra esta idea.

Ejemplo 2.7.5 Transformación aditiva

Considere un conjunto simple de datos ficticios, codificados restando 20 a cada observación. Las observaciones originales y transformadas se muestran en la Tabla 2.7.1.

Tabla 2.7.1 Efecto de una transformación aditiva				
	Observaciones originales (y)	Desviaciones ($y_i - \bar{y}$)	Observaciones transformadas (y')	Desviaciones ($y'_i - \bar{y}$)
	25	− 1	5	− 1
	26	0	6	0
	28	2	8	2
	25	− 1	5	− 1
Media	26		6	

La DT de las observaciones originales es

$$s = \sqrt{\frac{(-1)^2 + (0)^2 + (2)^2 + (-1)^2}{3}} = 1,4$$

Como las desviaciones no son afectadas por la transformación, la DT de las observaciones transformadas es la misma:

$$s' = 1,4$$

Una transformación aditiva desplaza el histograma de una distribución y lo mueve a la izquierda o a la derecha en la recta numérica. La forma del histograma no cambia y las desviaciones tampoco cambian, por lo que la DT no se modifica. Una transformación multiplicativa, por otra parte, ensancha o estrecha la distribución, por lo que la DT aumenta o disminuye en consecuencia.

Otros estadísticos Bajo transformaciones lineales, otras medidas de tendencia central (por ejemplo, la mediana) cambian como $\bar{y}$, y otras medidas de dispersión (por ejemplo, el intervalo intercuartílico), cambian como s . Los cuartiles cambian como $\bar{y}$.

Transformaciones no lineales

Algunas veces los datos se reexpresan de una forma no lineal. Algunos ejemplos de transformaciones no lineales son

$$Y' = \sqrt{Y}$$

$$Y' = \log(Y)$$

$$Y' = \frac{1}{Y}$$

$$Y' = Y^2$$

Esas transformaciones se denominan «no lineales» porque una gráfica de Y' con respecto a Y sería una curva en vez de una recta. Los computadores facilitan el uso de transformaciones no lineales. La transformación logarítmica es especialmente común en biología debido a que muchas relaciones importantes se pueden expresar de forma sencilla en función de logaritmos. Por ejemplo, hay una fase en el crecimiento de una colonia bacteriana en la que el logaritmo del tamaño de la colonia crece a una velocidad constante con el tiempo. [Nótese que los logaritmos se utilizan en algunas escalas familiares de medida, como la medida del pH o de la magnitud de terremotos (escala de Richter)].

Las transformaciones no lineales pueden afectar a los datos de forma compleja. Por ejemplo, la media puede no cambiar «naturalmente» bajo una transformación logarítmica. El logaritmo de la media *no* es lo mismo que la media de los logaritmos. Además, las transformaciones no lineales (a diferencia de las lineales) sí cambian la forma esencial de una distribución de frecuencias.

En capítulos posteriores veremos que si la distribución está sesgada hacia la derecha, como en el caso de la distribución del tiempo de canto de grillos que se muestra en la Figura 2.7.2, entonces puede ser conveniente aplicar una transformación que haga la distribución más simétrica, haciendo disminuir la cola derecha. Utilizando la transformación $Y' = \sqrt{Y}$ disminuye la cola derecha la distribución y aumenta su cola izquierda. La transformación $Y' = \log(Y)$ produce el mismo efecto pero de forma más severa que $\sqrt{Y}$. El ejemplo siguiente muestra el efecto de estas transformaciones.

Ejemplo 2.7.6 Tiempos de canto de grillos

La Figura 2.7.2(a) muestra la distribución de los datos de tiempos de canto de grillos de la Tabla 2.3.1. Si se transforman estos datos tomando raíces cuadradas, los datos transformados tienen la distribución que se muestra en la Figura 2.7.2(b). Tomando logaritmos (en base 10) resulta la distribución que se muestra en la Figura 2.7.2(c). Nótese que las transformaciones tienen el efecto de hacer disminuir la cola superior irregular «encogiéndola» los valores del extremo inferior de la distribución original.

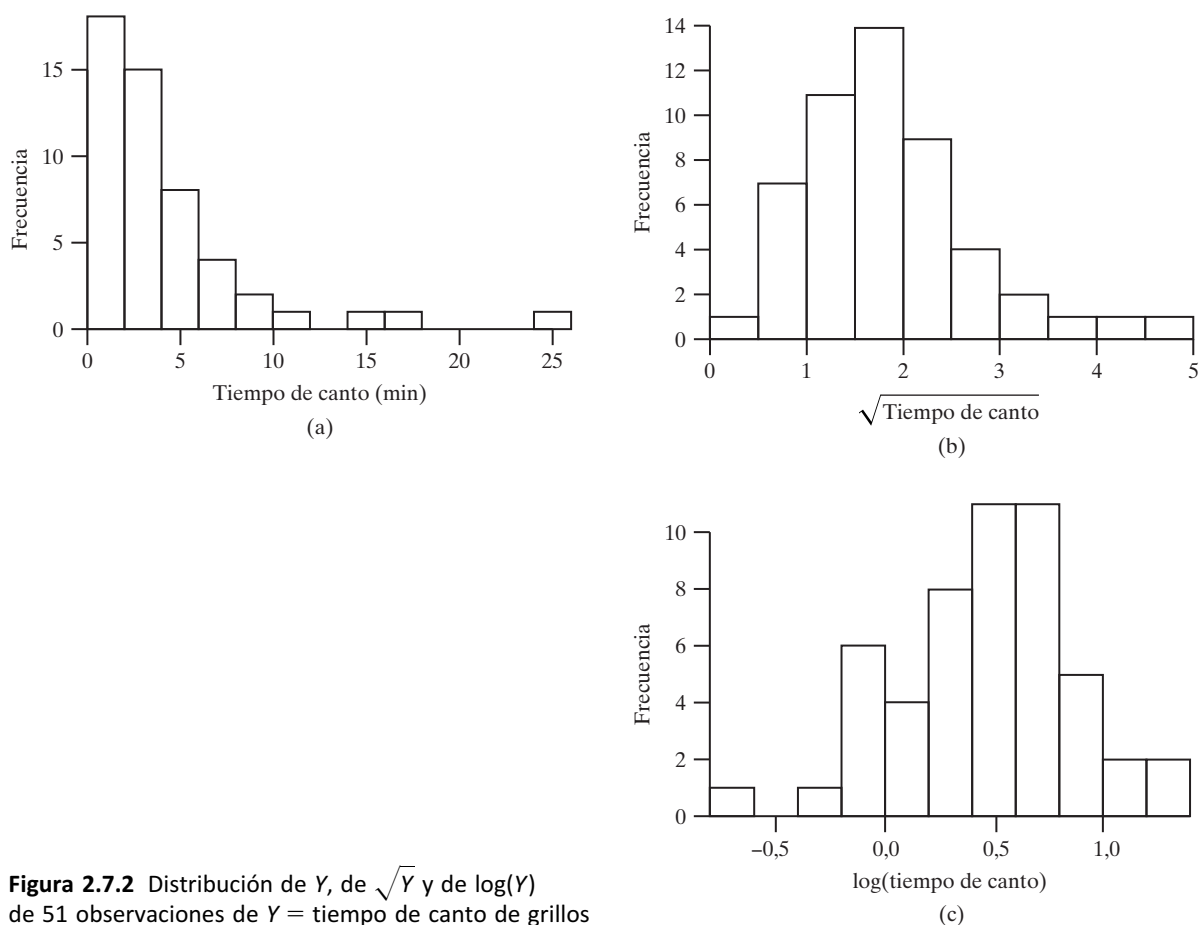


Figura 2.7.2 Distribución de Y , de $\sqrt{Y}$ y de $\log(Y)$ de 51 observaciones de $Y =$ tiempo de canto de grillos

Ejercicios 2.7.1-2.7.6

2.7.1 Una bióloga realizó una cierta medida del pH en 24 ranas. Algunos valores típicos fueron⁴⁴

7,43, 7,16, 7,51, ...

La bióloga calculó una media de 7,373 y la desviación típica de 0,129 de las medidas originales de pH. Seguidamente, transformó los datos restando siete a cada observación y después multiplicándolas por 100. Por ejemplo, 7,43 se transformó en 43. Los datos transformados son

43, 16, 51, ...

¿Cuáles son la media y la desviación típica de los datos transformados?

2.7.2 La media y la DT de un conjunto de 47 medidas de temperatura corporal fueron las siguientes⁴⁵:

$\bar{y} = 36,497^\circ\text{C}$ $s = 0,172^\circ\text{C}$

Si las 47 medidas se transforman a $^\circ\text{F}$,

(a) ¿Cuáles serían las nuevas media y DT?

(b) ¿Cuál sería el nuevo coeficiente de variación?

2.7.3 Un investigador midió las ganancias diarias medias (en kg/día) de 20 reses. Algunos valores típicos fueron⁴⁶

1,39, 1,57, 1,44, ...

La media de los datos fue de 1,461 y la desviación típica fue 0,178.

(a) Exprese la media y la desviación típica en lb/día (*Sugerencia*: 1 kg = 2,20 lb).

(b) Calcule el coeficiente de variación cuando los datos expresan (i) kg/día; (ii) lb/día.

2.7.4 Considere los datos del Ejercicio 2.7.3. La media y la DT fueron de 1,461 y 0,178. Supongamos que se transformarán los datos de

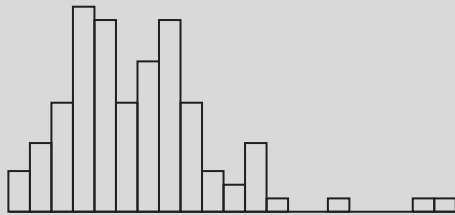
1,39, 1,57, 1,44, ...

por tanto,

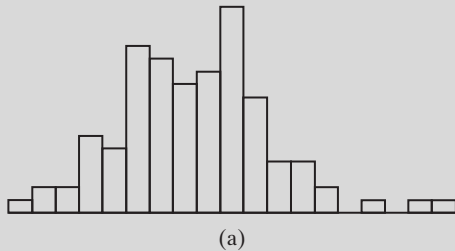
39, 57, 44, ...

¿Cuáles serían la media y la desviación típica de los datos transformados?

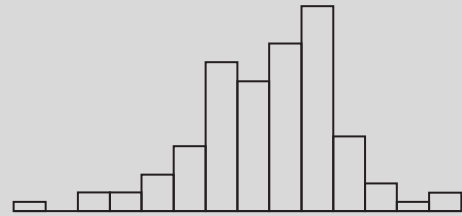
2.7.5 El siguiente histograma muestra la distribución de una muestra de datos:



Uno de los siguientes histograma es el resultado de aplicar una transformación de raíz cuadrada y el otro es el resultado de aplicar una transformación de logaritmo. ¿Cuál es cuál? ¿Cómo lo sabe?



(a)



(b)

2.7.6 (Problema para computador) El archivo 'Exer2.7.6.csv' se incluye en el disco de datos que se proporciona con este libro. Este archivo contiene 36 observaciones del número de segmentos de ramificación de dendritas de nueve células nerviosas tomadas de los cerebros de cerdos de guinea recién nacidos. (Estos datos se utilizaron en el Ejercicio 2.2.4). Abra el archivo e introduzca los datos en un paquete estadístico. Realice un histograma de los datos, que está sesgado hacia la derecha. Considere ahora las siguientes posibles transformaciones: $\sqrt{Y}$, $\log(Y)$ y $1/\sqrt{Y}$. ¿Cuál de estas transformaciones se comporta mejor para conseguir el objetivo de hacer que la distribución resultante sea razonablemente simétrica?

2.8 Inferencia estadística

La descripción de un conjunto de datos es algunas veces de interés en sí misma. Sin embargo, en general, los investigadores desearán generalizar, extender las conclusiones más allá del alcance limitado del grupo particular de animales, plantas, u otras unidades que han sido realmente observadas. La teoría estadística proporciona bases racionales para este proceso de generalización, a partir del modelo de muestra aleatoria de la Sección 1.3 y teniendo en cuenta la variabilidad de los datos. La idea clave del planteamiento estadístico es ver los datos concretos en un estudio como una muestra de una población mayor. La población es el foco real del interés científico y/o práctico. El siguiente ejemplo ilustra esta idea.

Ejemplo 2.8.1 Grupos sanguíneos

En un estudio inicial sobre el sistema ABO de clasificación de grupos sanguíneos, unos investigadores determinaron los grupos sanguíneos de 3.696 personas en Inglaterra. Los resultados se muestran en la Tabla 2.8.1⁴⁷.

Estos datos no se obtuvieron con el objetivo de saber los grupos sanguíneos de esas 3.696 personas en particular. Se obtuvieron por su valor científico como fuente de información sobre la distribución de los grupos sanguíneos en una población mayor. Por ejemplo, se podría suponer que la distribución de grupos sanguíneos de toda la población inglesa debería parecerse a la distribución de estas 3.696 personas. En particular, la frecuencia relativa observada del tipo sanguíneo A fue

$$\frac{1.634}{3.696} \quad \text{o} \quad 44 \% \text{ tipo A}$$

Tabla 2.8.1 Grupos sanguíneos de 3.696 personas	
Grupo sanguíneo	Frecuencia
A	1.634
B	327
AB	119
O	1.616
Total	3.696

A partir de esto, podría concluirse que aproximadamente el 44 % de las personas de Inglaterra tienen grupo sanguíneo A. ■

El proceso de obtener conclusiones sobre una población basadas en las observaciones de una muestra de dicha población se denomina **inferencia estadística**. Por ejemplo, en el Ejemplo 2.8.1, la conclusión de que aproximadamente el 44 % de las personas de Inglaterra tienen grupo sanguíneo A sería una inferencia estadística. La inferencia se muestra de forma esquemática en la Figura 2.8.1. Por supuesto, esa inferencia podría ser completamente errónea (quizá las 3.696 personas no son en absoluto representativas de la población inglesa en general). Nos podrían preocupar dos posibles fuentes de dificultad: (1) las 3.696 personas podrían haber sido seleccionadas de una forma que fuera sistemáticamente sesgada hacia (o contra) las personas de tipo A, y (2) el número de personas examinadas podría haber sido demasiado pequeño como para permitir la generalización a una población de muchos millones. En general, resulta que el hecho de que el tamaño de la población sea de millones *no* es un problema, pero el sesgo en la forma de la selección de las personas es un aspecto importante.

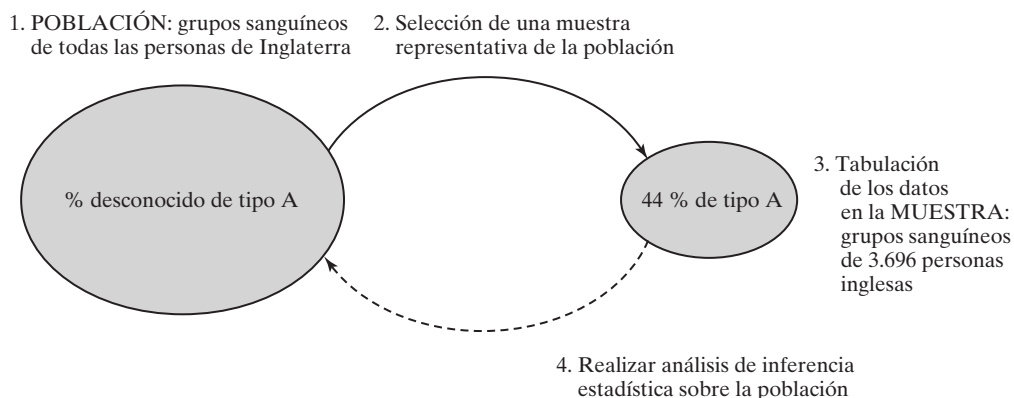


Figura 2.8.1 Representación esquemática de la inferencia de una muestra a la población respecto a la prevalencia del grupo sanguíneo A

Al realizar un inferencia estadística, esperamos que la muestra represente muy ajustadamente a la población, es decir, que la muestra sea *representativa* de la población. En la Sección 1.3 vimos cómo los errores de muestreo y los errores no debidos al muestreo pueden producir muestras más representativas. Sin embargo, incluso en ausencia de sesgo debemos preguntarnos sobre la verosimilitud de que una muestra particular proporcione una buena representación de la población. La pregunta importante es: *¿cuál es la verosimilitud de que una muestra sea representativa (de la población)?* En el Capítulo 5 veremos cómo la teoría estadística puede ayudarnos a responder a esta pregunta.

Especificación de la población

En la Sección 1.3 recalcamos que la obtención de los individuos que forman una muestra debería ser representativa de la población. De hecho, este requisito es un poco más fuerte de lo que es realmente necesario. En definitiva, lo que importa es que las medidas que tenemos de la variable de interés sean representativas de los valores presentes en la población. A continuación se proporciona un ejemplo de un caso en el que los miembros de la población podrían no ser representativos de dicha población, pero se podría argumentar que las medidas tomadas de esta muestra podrían verse como representativas de la población más grande.

Ejemplo 2.8.2 Grupos sanguíneos

¿Cómo se escogieron realmente las 3.696 personas inglesas del Ejemplo 2.8.1? Del artículo original parece que fue un «muestreo de conveniencia», es decir, amigos de los investigadores, empleados, y diversas fuentes no especificadas. Hay poca base para pensar que las propias *personas* sean representativas de la población inglesa completa. Sin embargo, se podría argumentar que sus grupos sanguíneos sí podrían ser (más o menos) representativos de la población. El argumento sería que los sesgos introducidos en la selección de esas personas concretas probablemente no estarían relacionados con el grupo sanguíneo. [No obstante, una objeción a este argumento podría basarse en la raza. Por ejemplo, la distribución racial de la muestra podría ser sustancialmente diferente de la distribución racial de Inglaterra (la población) y existen diferencias conocidas en las distribuciones de los grupos sanguíneos entre razas]. El argumento de la representatividad sería mucho menos convincente si la variable observada fuera la presión sanguínea en vez del grupo sanguíneo. Se sabe que la presión sanguínea tiende a aumentar con la edad, y el procedimiento de selección estuvo indudablemente sesgado en contra de ciertos grupos de edad (por ejemplo, las personas mayores). ■

Como muestra el Ejemplo 2.8.2, el que sea razonable pensar que las medidas obtenidas de una muestra sean representativas de las medidas de una población no depende solo en cómo las unidades observacionales (en este caso personas) han sido elegidas, sino también en la variable que ha sido observada. Idealmente, siempre deberíamos trabajar con muestras aleatorias, pero hemos visto que en algunos casos no es posible o conveniente obtener muestras aleatorias. Sin embargo, centrando nuestra atención en las propias medidas en vez de en los individuos de las que proceden, a menudo podemos plantear un argumento sobre la generalización (o no generalización) de nuestros resultados a una población mayor. Podemos hacer esto viendo la población como observaciones de un conjunto de valores de un proceso de medida, en vez de personas u otras unidades observacionales. A continuación sigue otro ejemplo.

Ejemplo 2.8.3 Alcohol y MOPEG

El compuesto bioquímico MOPEG tiene un papel en la función cerebral. Siete voluntarios masculinos con buena salud participaron en un estudio para determinar si beber alcohol podría elevar la concentración de MOPEG en el fluido cerebroespinal. La concentración de MOPEG se midió dos veces en cada hombre (una vez al comienzo del experimento y otra vez después de haber bebido 80 g de etanol). Los resultados (en pmol/ml) se muestran en la Tabla 2.8.2⁴⁸.

Centremos nuestra atención en la columna que está más a la derecha, que muestra el cambio en la concentración de MOPEG (es decir, la diferencia entre las medidas «después» y «antes»). Al pensar que estos valores son una muestra de una población, necesitamos especificar todos los detalles de las condiciones experimentales (cómo se obtuvieron los especímenes de fluido cerebroespinal, el momento exacto de las medidas y del consumo de alcohol, etc.), así como todas las características relevantes de los propios voluntarios. Por tanto, la definición de la población debería ser algo como esto:

Población Cambio en la concentración cerebroespinal de MOPEG en hombres jóvenes saludables cuando se mide antes y después de beber 80 g de etanol, estando ambas medidas hechas a las 8 de la mañana, ... (aquí se especificarían otras condiciones experimentales relevantes).

Tabla 2.8.2 Efectos del alcohol sobre el MOPEG			
Voluntario	Concentración de MOPEG		
	Antes	Después	Cambio
1	46	56	10
2	47	52	5
3	41	47	6
4	45	48	3
5	37	37	0
6	48	51	3
7	58	62	4

No existe una única definición «correcta» de población en un experimento como este. Un científico que lea un informe del experimento podría encontrar esta definición demasiado exhaustiva (por ejemplo, quizá no importa que las medidas de los voluntarios se tomaran a las 8 de la mañana), o demasiado general. Podría entonces utilizar su conocimiento sobre el alcohol y la química del cerebro para formular su propia definición, y utilizar esa definición con una base para interpretar estas siete observaciones. ■

Descripción de una población

Como las observaciones se realizan solamente sobre una muestra, casi nunca se conocen de forma exacta las características de las poblaciones biológicas. Generalmente, nuestro conocimiento de una característica de una población proviene de una muestra. En lenguaje estadístico, podemos decir que la característica de la muestra es una estimación de la correspondiente característica de la población. Por tanto, la estimación es un tipo de inferencia estadística.

De la misma forma que cada muestra tiene una distribución, una media y una DT, podemos imaginar una distribución de la población, una media de la población y una DT de la población. Para presentar los conceptos de la inferencia de una muestra a una población, necesitamos un lenguaje para describir la población. Este lenguaje es paralelo al lenguaje que describen la muestra. Una característica de la muestra se denomina **estadístico**; una característica de la población se denomina **parámetro**.

Proporciones

Dada una variable categórica, podemos describir una población indicando simplemente la proporción o frecuencia relativa de la población en cada categoría. Se presenta a continuación un ejemplo sencillo.

Ejemplo 2.8.4 Plantas de avena

En una cierta población de plantas de avena la resistencia a la enfermedad de la roya de la hoja se distribuye como se muestra en la Tabla 2.8.3⁴⁹. ■

Tabla 2.8.3 Resistencia a la enfermedad en plantas de avena	
Resistencia	Proporción de plantas
Resistentes	0,47
Resistencia intermedia	0,43
Susceptibles	0,10
Total	1,00

Observación La población descrita en el Ejemplo 2.8.4 es realista, pero no es una población real específica. Las proporciones exactas para cualquier población real no son conocidas. Por razones similares, utilizaremos poblaciones ficticias aunque realistas en algunos otros ejemplos, aquí y en los Capítulos 3, 4 y 5.

En el caso de datos categóricos, la proporción muestral de una categoría es una estimación de la correspondiente proporción poblacional. Dado que esas proporciones no son necesariamente iguales, es esencial tener una notación para distinguirlas. Indicaremos como p a la proporción poblacional de una categoría y como $\hat{p}$ (léase «p-gorro») a la proporción muestral:

p = proporción poblacional

$\hat{p}$ = proporción muestral

El símbolo « $\hat{}$ » se puede interpretar como «estimador de». Por tanto,

$\hat{p}$ es un estimador de p

Ilustraremos esta notación con un ejemplo.

Ejemplo 2.8.5 Cáncer de pulmón

Once pacientes de adenocarcinoma (un tipo de cáncer de pulmón) fueron tratados con el agente quimioterápico Mytomicyn. Tres de los pacientes mostraron una respuesta positiva (definida como una reducción del tumor de al menos el 50 %)⁵⁰. Supongamos que definimos la población de este estudio como «respuestas de todos los pacientes de adenocarcinoma». Entonces podemos representar las proporciones muestral y poblacional de la categoría «respuesta positiva» de la siguiente forma:

p = Proporción de respuestas positivas entre todos los pacientes de adenocarcinoma

$\hat{p}$ = Proporción de respuestas positivas entre los 11 pacientes de este estudio

$$\hat{p} = \frac{3}{11} = 0,27$$

Nótese que p es desconocido, y $\hat{p}$, que es conocido, es un estimador de p . ■

Debemos recalcar que un «estimador», tal como estamos utilizando el término, puede ser o no ser un *buen* estimador. Por ejemplo, el estimador $\hat{p}$ en el Ejemplo 2.8.5 se basa en muy pocos pacientes. Los estimadores basados en un pequeño número de observaciones están sujetos a una incertidumbre considerable. Por supuesto, la cuestión de si un procedimiento de estimación es bueno o pobre es importante, y en capítulos posteriores mostraremos cómo se puede responder a esta pregunta.

Otras medidas descriptivas

Si la variable observada es cuantitativa, se pueden considerar otras medidas descriptivas además de las proporciones: la mediana, los cuartiles, la DT, etc. Todas esas cantidades se pueden calcular dada una muestra de datos, y cada una es un estimador de su correspondiente análogo poblacional. Por ejemplo, la mediana muestral es una estimación de la mediana poblacional. En capítulos posteriores nos centraremos principalmente en la media y en la DT, y por lo tanto necesitamos una notación especial para la media y la DT poblacionales. **La media poblacional se denotará como μ (mu) y la DT poblacional se denotará como σ (sigma)**. Se pueden definir como sigue para una variable cuantitativa Y :

μ = valor del promedio poblacional de Y

$$\sigma = \sqrt{\text{valor del promedio poblacional de } (Y - \mu)^2}$$

El siguiente ejemplo ilustra esta notación.

Ejemplo 2.8.6 Hojas de tabaco

Un agrónomo contó el número de hojas en cada una de 150 plantas de tabaco de la misma variedad (Habana). Los resultados se muestran en la Tabla 2.8.4⁵¹.

La media muestral es

$$\bar{y} = 19,78 = \text{número medio de hojas en las 150 plantas}$$

Tabla 2.8.4 Número de hojas en plantas de tabaco	
Número de hojas	Frecuencia (número de plantas)
17	3
18	22
19	44
20	42
21	22
22	10
23	6
24	1
Total	150

La media poblacional es

$$\mu = \text{número medio de hojas de plantas de tabaco Habana que han crecido bajo las mismas condiciones}$$

Como no conocemos μ , podemos utilizar $\bar{y} = 19,78$ como una estimación de μ . La DT muestral es

$$s = 1,38 = \text{DT del número de hojas en las 150 plantas}$$

La DT poblacional es

$$\sigma = \text{DT del número de hojas de plantas de tabaco Habana que han crecido bajo las mismas condiciones}$$

Como no conocemos σ , podemos utilizar $s = 1,38$ como una estimación de σ^* . ■

2.9 Perspectiva

En este capítulo hemos considerado varias formas de describir un conjunto de datos. Hemos presentado también las características de una muestra como estimadores de las correspondientes características de una población definida adecuadamente.

Parámetros y estadísticos

Algunas características de la distribución (por ejemplo, la media) se pueden representar mediante un único número, mientras que otras (por ejemplo, la forma) no. Hemos indicado que una medida numérica que describe la muestra se denomina estadístico. De forma correspondiente, una medida numérica que describe una población se denomina parámetro. Para el caso de las medidas numéricas más importantes, hemos definido notaciones que permiten distinguir entre el estadístico y el parámetro. Estas notaciones se resumen en la Tabla 2.9.1 para que sirvan de referencia.

Tabla 2.9.1 Notación de algunos estadísticos y parámetros importantes		
Medida	Valor muestral (estadístico)	Valor poblacional (parámetro)
Proporción	$\hat{p}$	p
Media	$\bar{y}$	μ
Desviación típica	s	σ

* Podemos preguntarnos por qué utilizamos $\bar{y}$ y s en lugar de $\hat{\mu}$ y $\hat{\sigma}$. Una respuesta de la tradición. Otra respuesta es que como « $\hat{\cdot}$ » significa estimador, se podrían tener otros estimadores en mente.

Una mirada hacia delante

Es natural ver una característica muestral (por ejemplo, $\bar{y}$) con un estimador de la correspondiente característica poblacional (por ejemplo, μ). Pero adoptando esta perspectiva, hay que estar alerta contra optimismos injustificados. Por supuesto, si la muestra fuera perfectamente representativa de la población, entonces el estimador sería perfectamente exacto. Pero esto hace surgir la pregunta central: ¿cómo de verosímil es que la muestra sea representativa (de la población)? La intuición sugiere que, si las unidades observacionales se seleccionan apropiadamente, entonces la muestra debería ser más o menos representativa de la población. La intuición también sugiere que muestras grandes deberían tender a ser más representativas que muestras pequeñas. Estas impresiones son básicamente correctas, pero son demasiado vagas como para proporcionar una guía práctica para investigar en las Ciencias de la Vida. Las cuestiones prácticas que necesitamos responder son

1. ¿Cómo puede juzgar un investigador si una muestra se puede ver como «más o menos» representativa de una población?
2. ¿Cómo puede un investigador cuantificar «más o menos» en un caso concreto?

En la Sección 1.3 describimos un modelo teórico de probabilidad basado en el muestreo aleatorio que proporciona un marco para juzgar la pregunta (1), y en el Capítulo 6 veremos cómo este modelo puede proporcionar una respuesta concreta a la pregunta (2). Concretamente, en el Capítulo 6 veremos cómo analizar un conjunto de datos de forma que se pueda cuantificar la cercanía de los estimadores de la media muestral ($\bar{y}$) a la media poblacional (μ). Pero antes de volver al análisis de datos del Capítulo 6, necesitamos presentar algunas bases en los Capítulos 3, 4 y 5. Los desarrollos de estos capítulos son un preludio esencial para entender las técnicas de inferencia estadística.

Ejercicios suplementarios 2.S.1-2.S.20

2.S.1 Una muestra de cuatro estudiantes tuvieron las siguientes alturas (en cm): 1,80, 1,82, 1,79, 1,76. Suponga que se añade al grupo un quinto estudiante. ¿Qué altura debería tener ese estudiante para hacer que la media del grupo fuera igual a 1,81?

2.S.2 Una botánica hizo crecer 15 plantas de pimiento en el mismo banco de un invernadero. Después de 21 días, midió la longitud total del tallo (cm) de cada planta, y obtuvo los siguientes valores⁵²:

12,4	12,2	13,4
10,9	12,2	12,1
11,8	13,5	12,0
14,1	12,7	13,2
12,6	11,9	13,1

- (a) Construya un diagrama de puntos de estos datos, y marque la posición de los cuartiles.
- (b) Calcule el intervalo intercuartílico.

2.S.3 En un estudio sobre el comportamiento de la mosca de la fruta *Drosophila melanogaster*, un biólogo midió, en moscas individuales, el tiempo total empleado en arreglarse las salas durante un periodo observación de seis minutos. Se presentan a continuación los tiempos (en segundos) de 20 moscas⁵³:

34	24	10	16	52
76	33	31	46	24
18	26	57	32	25
48	22	48	29	19

- (a) Determine la mediana y los cuartiles.
- (b) Determine el intervalo intercuartílico.
- (c) Construya un diagrama de caja (modificado) de los datos.

2.S.4 Para calibrar una curva estándar para ensayos sobre concentraciones de proteínas, un patólogo de plantas utilizó un espectrómetro para medir la absorción de la luz (en una longitud de onda de 50 nm) de una solución de proteína. El resultado de 27 repeticiones del ensayo sobre una solución estándar que contenía 60 μg de proteína por ml de agua fueron los siguientes⁵⁴:

0,111	0,115	0,115	0,110	0,099
0,121	0,107	0,107	0,100	0,110
0,106	0,116	0,098	0,116	0,108
0,098	0,120	0,123	0,124	0,122
0,116	0,130	0,114	0,100	0,123
0,119	0,107			

Construya una distribución de frecuencias y preséntela en forma de tabla y de histograma.

2.S.5 Refiriéndonos a los datos de absorción del Ejercicio 2.S.4.

- (a) Determine la mediana, los cuartiles y el intervalo intercuartílico.
- (b) ¿Cómo tendría que ser de grande una observación para ser un *outlier*?

2.S.6 El rango medio se define como el promedio del mínimo y el máximo de la distribución. ¿Es el rango medio un estadístico robusto? ¿Por qué o por qué no?

2.S.7 20 pacientes con epilepsia severa fueron observados durante ocho semanas. Se presentan a continuación los números de ataques mayores sufridos por cada paciente durante el periodo de observación⁵⁵:

5 0 9 6 0 0 5 0 6 1
5 0 0 0 0 7 0 0 4 7

- Determine la mediana del número de ataques.
- Determine el número medio de ataques.
- Construya un histograma de los datos. Marque en dicho histograma las posiciones de la media y de la mediana.
- ¿Qué característica de la distribución de frecuencias sugiere que ni la media ni la mediana resumen de forma significativa la experiencia de estos pacientes?

2.S.8 Calcule la desviación típica de cada una de las siguientes muestras ficticias:

- 11, 8, 4, 10, 7.
- 6, 0, -3, 2, 5.
- 23, 29, 24, 21, 23.

2.S.9 Para estudiar la distribución espacial de las larvas de escarabajo japonés en el suelo, unos investigadores dividieron una sección de 12×12 pies de un maizal en 144 cuadrados de un pie. Contaron el número de larvas Y en cada cuadrado, con los resultados que se muestran en la siguiente tabla⁵⁶.

Número de larvas	Frecuencia (número de cuadrados)
0	13
1	34
2	50
3	18
4	16
5	10
6	2
7	1
Total	144

- La media y desviación típica de Y son $\bar{y} = 2,23$ y $s = 1,47$. ¿Qué porcentaje de las observaciones están dentro de un intervalo de
 - 1 desviación típica alrededor de la media?
 - 2 desviaciones típicas alrededor de la media?
- Determine el número total de larvas en los 144 cuadrados. ¿Cómo se relaciona este número con $\bar{y}$?
- Determine la mediana de la distribución.

2.S.10 Una medida de la forma física es el consumo máximo de oxígeno, que es la velocidad máxima a la que una persona puede consumir oxígeno. Se utilizó un test basado en cinta de correr para determinar el máximo consumo de oxígeno de nueve universitarias antes y después de participar en un programa de 10 semanas de ejercicio enérgico. La tabla que sigue muestra las medidas antes y después del cambio (después-antes). Todos los valores están en ml de O_2 por kg de peso corporal⁵⁷.

Participante	Consumo máximo de oxígeno		
	Antes	Después	Cambio
1	48,6	38,8	-9,8
2	38,0	40,7	2,7
3	31,2	32,0	0,8
4	45,5	45,4	-0,1
5	41,7	43,2	1,5
6	41,8	45,3	3,5
7	37,9	38,9	1,0
8	39,2	43,5	4,3
9	47,2	45,0	-2,2

Los cálculos siguientes se realizarán sobre el *cambio* en el consumo máximo de oxígeno (la columna más a la derecha).

- Calcule la media y la desviación típica.
- Determine la mediana.
- Elimine la participante 1 de los datos y repita los apartados (a) y (b). ¿Cuál de las medidas descriptivas muestran resistencia y cuál no?

2.S.11. Un anatomista veterinario investigó la disposición espacial de las células nerviosas en el intestino de un pony. Recogió un bloque de tejido de la pared intestinal, cortó el bloque en muchas secciones iguales, y contó el número de células nerviosas en cada una de 23 secciones seleccionadas aleatoriamente. Los resultados fueron los siguientes⁵⁸.

35 19 33 34 17 26 16 40
28 30 23 12 27 33 22 31
28 28 35 23 23 19 29

- Determine la media, los cuartiles y el intervalo intercuartílico.
- Construya un diagrama de caja de los datos.

2.S.12. En el Ejercicio 2.S.11 se pide el diagrama de caja de los datos de células nerviosas. ¿Permite este gráfico sostener la afirmación de que los datos vienen de una distribución razonablemente simétrica?

2.S.13 Un genetista contó el número de cerdas en una cierta región del abdomen de la mosca de la fruta *Drosophila melanogaster*. Los resultados de 119 individuos fueron los que se muestran en la tabla⁵⁹.

Número de cerdas	Número de moscas	Número de cerdas	Número de moscas
29	1	38	18
30	0	39	13
31	1	40	10
32	2	41	15
33	2	42	10
34	6	43	2
35	9	44	2
36	11	45	3
37	12	46	2

- Calcule la mediana del número de cerdas.
- Calcule los cuartiles primero y tercero de la muestra.
- Realice un diagrama de caja de los datos.
- La media de la muestra es 38,45 y la desviación típica es 3,20. ¿Qué porcentaje de las observaciones está dentro de un intervalo de desviación típica alrededor de la media?

2.S.14 Se piensa que el monóxido de carbono de los cigarrillos puede producir daños en el feto de una mujer embarazada cuando fuma. En un estudio de esta hipótesis se extrajo sangre de mujeres embarazadas antes y después de fumar un cigarrillo. Se hicieron medidas del porcentaje de hemoglobina en la sangre combinada con el monóxido de carbono en forma de carboxihemoglobina (COHb). Los resultados de las mujeres se muestran en la tabla⁶⁰.

Sujeto	COHB en sangre (%)		
	Antes	Después	Incremento
1	1,2	7,6	6,4
2	1,4	4,0	2,6
3	1,5	5,0	3,5
4	2,4	6,3	3,9
5	3,6	5,8	2,2
6	0,5	6,0	5,5
7	2,0	6,4	4,4
8	1,5	5,0	3,5
9	1,0	4,2	3,2
10	1,7	5,2	3,5

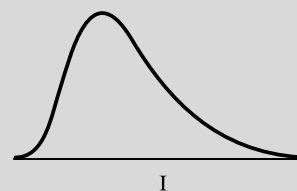
- Calcule la media y la desviación típica del *incremento* en COHb.
- Calcule la COHb media antes y la COHb media después. ¿Es el incremento medio igual a la media de los incrementos?
- Determine la mediana del incremento en COHb.
- Repita el apartado (c) para las medidas antes y para las medidas después. ¿Es la mediana del incremento igual al incremento de las medianas?

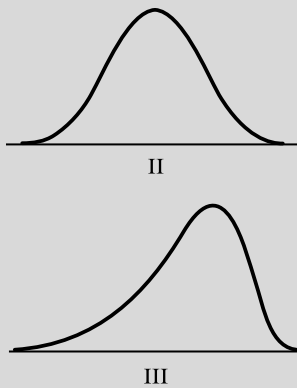
2.S.15 (Problema para computador) Un investigador médico de la India extrajo especímenes de sangre de 31 niños jóvenes, todos ellos infectados con malaria. Los siguientes datos, presentados en orden creciente, son los números de parásitos de malaria encontrados en 1 ml de sangre de cada niño⁶¹.

100	140	140	271	400	435	455	770
826	1.400	1.540	1.640	1.920	2.280	2.340	3.672
4.914	6.160	6.560	6.741	7.609	8.547	9.560	10.516
14.960	16.855	18.600	22.995	29.800	83.200	134.232	

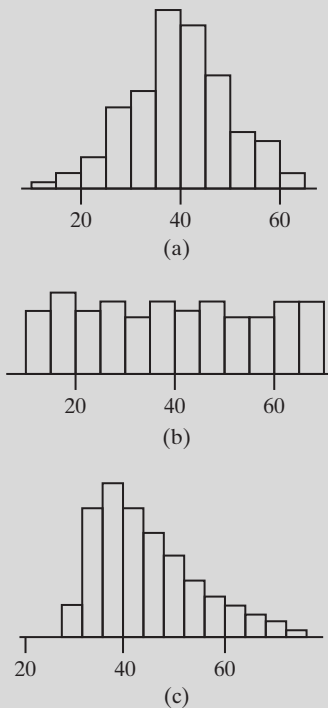
- Construya una distribución de frecuencias de los datos utilizando una anchura de clase de 10.000. Presente la distribución en forma de histograma.
- Transforme los datos tomando el logaritmo (en base 10) de cada observación. Construya una distribución de frecuencias de los datos transformados y preséntela en forma de histograma. ¿Cómo afecta la transformación logarítmica a la forma de la distribución de frecuencias?
- Determine la media de los datos originales y la media de los datos transformados con el logaritmo. ¿Es la media de los logaritmos igual al logaritmo de la media?
- Determine la mediana de los datos originales y la mediana de los datos transformados con el logaritmo. ¿Es la mediana de los logaritmos igual al logaritmo de la mediana?

2.S.16 Se midió la cantidad de precipitación, medida en pulgadas, durante el mes de junio en Cleveland, Ohio, durante 41 años⁶². Los valores tenían un mínimo de 1,2, una media de 3,6 y una desviación típica de 1,6. ¿Cuál de los siguientes histogramas puede ser un histograma aproximado de los datos? ¿Cómo lo sabe?





2.S.17 Los siguientes histogramas (a), (b) y (c) muestran tres distribuciones.

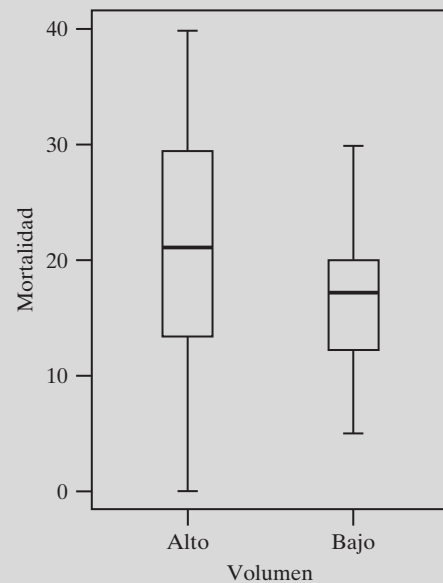


La siguiente salida de computador muestra la media, mediana y desviación típica de las tres distribuciones, más la media, mediana y desviación típica de una cuarta distribución. Haga corresponder los histogramas con los estadísticos. Explique su razonamiento. (Un conjunto de estadísticos no se utilizará).

1. Count	100	2. Count	100
Mean	41.3522	Mean	39.6761
Median	39.5585	Median	39.5377
StdDev	13.0136	StdDev	10.0476

3. Count	100	4. Count	100
Mean	37.7522	Mean	39.6493
Median	39.5585	Median	39.5448
StdDev	13.0136	StdDev	17.5126

2.S.18 Los siguientes diagramas de caja muestran las tasas de mortalidad (muertes al año cada 100 pacientes) de pacientes de trasplante de corazón en varios hospitales. Los hospitales de bajo volumen son aquellos que realizan entre 5 y 9 trasplantes al año. Los hospitales de alto volumen realizan 10 o más trasplantes al año⁶³. Describa las distribuciones, poniendo especial atención a cómo se comparan entre sí. Asegúrese de tener en cuenta la forma, el centro y la dispersión de cada distribución.

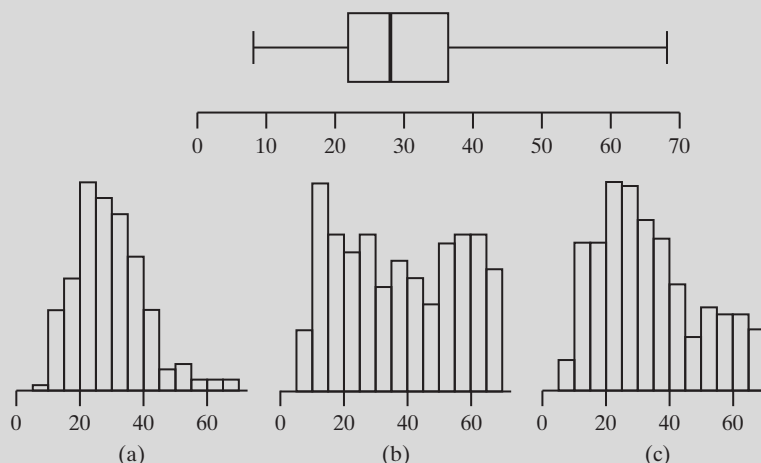


2.S.19 (Problema para computador) Unos médicos midieron la concentración de calcio (nM) en muestra de sangre de 38 personas saludables. Los datos se muestran a continuación⁶⁴.

95	110	135	120	88	125
112	100	130	107	86	130
122	122	127	107	107	107
88	126	125	112	78	115
78	102	103	93	88	110
104	122	112	80	121	126
90	96				

Calcule medidas apropiadas de tendencia central y de dispersión de la distribución. Describa la forma de la distribución y cualquier característica inusual de los datos.

2.S.20 El diagrama de caja siguiente muestra los mismos datos que en uno de los tres histogramas. ¿Qué histograma se corresponde con el diagrama de caja? Explique su respuesta.



Notas

1. Stewart, R. N. y Arisumi, T. (1966). Genetic and histogenic determination of pink bract color in poinsettia. *Journal of Heredity* **57**, 217-220.
2. Haselgrove, C., Straker, L., Smith, A., O'Sullivan, P., Perry, M. y Sloan, N. (2008). Perceived school bag load, duration of carriage, and method of transport to school are associated with spinal pain in adolescents: An observational study. *Australian Journal of Physiotherapy* **54**, 193-200.
3. Datos obtenidos online de *The World Factbook* produced by the CIA at www.cia.gov/library/publications/the-world-factbook/
4. Datos no publicados cortesía de C. M. Coxand K. J. Drewry.
5. Datos no publicados cortesía de W. F. Jacobson.
6. Datos no publicados recogidos en Oberlin College por J. Witmer.
7. Knoll, A. E. y Barghoorn, E. S. (1977). Archean microfossils showing cell division from the Swaziland system of South Africa. *Science* **198**, 396-398.
8. Nurse, C. A. (1981). Interactions between dissociated rat sympathetic neurons and skeletal muscle cells developing in cell culture. II. Synaptic mechanisms. *Developmental Biology* **88**, 71-79.
9. Topinard, P. (1888). Lepoids de l'encephale d'après les registres de Paul Broca. *Memoires Societe d'Anthropologie Paris*, 2nd series **3**, 1-41. Los datos que se muestran son un subconjunto de los datos publicados por Topinard.
10. Johannsen, W. (1903). *Ueber Erblichkeit in Populationen und in reinen Linien*. Jena: G. Fischer. Data reproduced in Strickberger, M. W. (1976). *Genetics*, New York: Macmillan, p. 277; y Peters, J. A. (ed.) (1959). *Classic Papers in Genetics*, Englewood Cliffs, N. J.: Prentice Hall, p. 23.
11. Datos no publicados cortesía de W. F. Jacobson.
12. Simpson, G. G., Roe, A. y Lewontin, R. C. (1960). *Quantitative Zoology*. New York: Harcourt, Brace, p. 51.
13. Adaptado de Potkin, S. G., Cannon, H. F., Murphy, D. L. y Wyatt, R. J. (1978). Are paranoid schizophrenics biologically different from other schizophrenics? *New England Journal of Medicine* **298**, 61-66. Los datos que se dan son aproximados, se han reconstruido del histograma y de la información resumen dada por Potkin *et al.* Reimpreso con permiso del *New England Journal of Medicine*.
14. Peters, H. G. y Bademan, H. (1963). The form and growth of stellate cells in the cortex of the guinea-pig. *Journal of Anatomy* (London) **97**, 111-117.
15. Datos cortesía de R. F. Jones, Indiana State Dairy Association, Inc.
16. Datos no publicados cortesía de D. J. Honor y W. A. Vestre.
17. Hepp, J., Buck, C. y Catalano, J. (2007). Analysis of three corn hybrids to determine if more expensive varieties are worth their cost. Manuscrito no publicado, Oberlin College.

18. Bruce, D., Harvey, D., Hamerton, A. E. y Bruce, L. (1913). Morphology of various strains of the trypanosome causing disease in man in Nyasaland. I. The human strain. *Proceedings of the Royal Society of London, Series B* **86**, 285-302. Véase también Pearson, K. (1914). On the probability that two independent distributions of frequency are really samples of the same population, with reference to recent work on the identity of trypanosome strains. *Biometrika* **10**, 85-143.
19. Shields, D. R. (1981). The influence of niacin supplementation on growing ruminants and *in vivo* and *in vitro* rumen parameters. Ph. D. thesis, Purdue University. Datos iniciales cortesía del autor y de D. K. Colby.
20. Gwynne, D. T. (1981). Sexual difference theory: Mormon crickets show role reversal in mate choice. *Science* **213**, 779-780. Copyright 1981 de la AAAS. Datos iniciales cortesía del autor.
21. Datos no publicados cortesía de M. A. Morse y G. P. Carlson.
22. Adaptado de Anderson, J. W., Story, L., Sieling, B., Chen, W. L., Petro, M. S. y Story, J. (1984). Hypocholesterolemic effects of oat-bran or bean intake for hypercholesterolemic men. *American Journal of Clinical Nutrition* **40**, 1146-1155. Había realmente 20 hombres en el estudio.
23. Datos no publicados cortesía de C. H. Noller.
24. Luria, S. F. y Delbruck, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491-511.
25. Datos ficticios pero realistas. Véase Roberts, J. (1975). Blood pressure of persons 18-74 years, United States, 1971-72. U. S. National Center for Health Statistics, *Vital and Health Statistics, Series 11*, n.º 150. Washington, D. C.: U. S. Department of Health, Education and Welfare.
26. Datos no publicados recogidos de una muestra de estudiantes del Oberlin College.
27. Datos no publicados cortesía de M. Kimmel.
28. Datos no publicados cortesía de F. Delgado.
29. Kitts, C., Moline, M., Schaffner, A., Samadpour, M., MacNiel, K. y Duffield, S. (2002). *Identifying the Sources of Escherichia coli Contamination in Crassostrea gigas from the Morro Bay Estuary*. Informe técnico para el National Estuary Program y el California Central Coast Regional Water Quality Control Board.
30. Kinghorn, A., Humphries, M., Outridge, P. y Chan H. M. (2008). Teeth as biomonitors of selenium concentrations in tissues of beluga whales (*Delphinapterus leucas*). *Science of the Total Environment* **402**, 43-50.
31. Govind, C. K. y Pearce, J. (1986). Differential reflex activity determines claw and closer muscle asymmetry in developing lobsters. *Science* **233**, 354-356. Copyright 1986 de la AAAS.
32. Adaptado de Gerdes, N. (2001). Morphological and life history variation in three populations of golden-mantled ground squirrels along a Pacific coast transect. Master's thesis California Polytechnic State University, San Luis Obispo.
33. Adaptado de Barclay, A. M. y Crawford, R. M. M. (1984). Seedling emergence in the rowan (*Sorbus aucuparia*) from an altitudinal gradient. *Journal of Ecology* **72**, 627-636. Reimpreso con permiso de Blackwell Scientific Publications Limited.
34. Datos ficticios pero realistas. Basado en Beyl, C. A. y Mitchell, C. A. (1977). Characterization of mechanical stress dwarfing in chrysanthemum. *Journal of the American Society for Horticultural Science* **102**, 591-594.
35. Basado en un subconjunto de los datos de Tuddenham, R. D. y Snyder, M. M. (1954). Physical growth of California boys and girls from birth to age 18. *Calif. Publ. Child Develop.* **1**, 183-364. Datos como se muestran en Weisberg, S. (1985). *Applied Linear Regression*, 2.ª ed. New York: Wiley, p. 57.
36. Nelson, L. A. (1980). Informe del the Indiana Beef Evaluation Program, Inc. Purdue University, West Lafayette, Indiana.
37. Datos recogidos por J. Witmer en un workshop sobre estadística en la Johns Hopkins University, julio de 1995.
38. Day, K. M., Patterson, F. L., Luetkemeier, O. W., Ohm, H. W., Polizotto, K., Roberts, J. J., Shaner, G. E., Huber, D. M., Finney, R. F., Foster, J. F. y Gallun, R. L. (1980). Performance and adaptation of small grains in Indiana. Station Bulletin No. 290. West Lafayette, Ind., Agricultural Experiment Station of Purdue University. Datos iniciales proporcionados por cortesía de W. F. Nyquist.
39. Tripepi, R. R. y Mitchell, C. A. (1984). Metabolic response of river birch and European birch roots to hypoxia. *Plant Physiology* **76**, 31-35. Datos iniciales proporcionados por cortesía de los autores.
40. Ogilvie, R. I., Macleod, S., Fernandez, P. y McCullough, W. (1974). Timolol in essential hypertension. En B. Magnani (ed.). *Beta-Adrenergic Blocking Agents in the Management of Hypertension and Angina Pectoris*. New York: Raven Press, pp. 31-43.
41. Datos no publicados cortesía de J. F. Nashy y J. E. Zabik.
42. Schall, J. J., Bennett, A. F. y Putnam, R. W. (1982). Lizards infected with malaria: Physiological and behavioral consequences. *Science* **217**, 1057-1059. Copyright 1982 de la AAAS. Datos iniciales proporcionados por cortesía de J. J. Schall.
43. Datos ficticios pero realistas. Cada observación es el promedio de varias medidas hechas a cada mujer en momentos diferentes. Véase Royston, J. P. y Abrams, R. M. (1980). An objective method for detecting the shift in basal body temperature in women. *Biometrics* **36**, 217-224.

44. Adaptado de los datos de Cicirelli, M. F., Robinson, K. R. y Smith, L. D. (1983). Internal pH of *Xenopus oocytes*: A study of the mechanism and role of pH changes during meiotic maturation. *Developmental Biology* **100**, 133-146.
45. Adaptado de Royston, J. P. y Abrams, R. M. (1980). An objective method for detecting the shift in basal body temperature in women. *Biometrics* **36**, 217-224.
46. Adaptado de los datos proporcionados por cortesía de L. A. Nelson.
47. Ikin, E. W., Prior, A. M., Race, R. R. y Taylor, G. L. (1939). The distribution of the A₁A₂BO blood groups in England. *Annals of Eugenics* (London) **9**, 409-411. Reimpreso con permiso de Cambridge University Press.
48. Borg, S., Kvande, H. y Sedvall, G. (1981). Central norepinephrine metabolism during alcohol intoxication in addicts and healthy volunteers. *Science* **213**, 1135-1137. Copyright 1981 de la AAAS. Datos iniciales cortesía de S. Borg.
49. Población ficticia pero realista. Adaptado de LeClerc, E. L., Leonard, W. H. y Clark, A. G. (1962). *Field Plot Technique*. Minneapolis: Burgess.
50. Selawry, O. S. (1974). The role of chemotherapy in the treatment of lung cancer. *Seminars in Oncology* **1**, n.º 3, 259-272.
51. Hayes, H. K., East, E. M. y Bernhart, E. G. (1913). *Connecticut Agricultural Experiment Station Bulletin* 176. Datos reproducidos en Strickberger, M. W. (1976). *Genetics*. New York: Macmillan, p. 288.
52. Datos no publicados cortesía de J. Y. Latimer y C. A. Mitchell.
53. Connolly, K. (1968). The social facilitation of preening behaviour in *Drosophila melanogaster*. *Animal Behaviour* **16**, 385-391.
54. Los resultados de ensayos similares se presentan en Pascholati, S. F. y Nicholson, R. L. (1983). *Helminthosporium maydis* suppresses expression of resistance to *Helminthosporium carbonum* in corn. *Phytopathologische Zeitschrift* **107**, 97-105. Datos no publicados cortesía de los investigadores.
55. Richens, A. y Ahmad, S. (1975). Controlled trial of valproate in severe epilepsy. *British Medical Journal* **4**, 255-256.
56. Fleming, W. E. y Baker, F. E. (1936). A method for estimating populations of larvae of the Japanese beetle in the field. *Journal of Agricultural Research* **53**, 319-331. Datos reproducidos en *Statistical Ecology*, vol. 1 (1971). University Park: Pennsylvania State University Press, p. 327.
57. Chiarotti, R. M. (1972). An investigation of the energy expenditure of women squash players. Master's thesis, Pennsylvania State University. Datos iniciales cortesía de R. M. Lyle (nee Chiarotti).
58. Mast, J. (1983). Innervation of the equine small intestine. Master's thesis, Purdue University. Datos iniciales cortesía del autor.
59. Datos ficticios pero realistas. Adaptados de datos presentados en Falconer, D. S. (1981). *Introduction to Quantitative Genetics*, 2.^a ed. New York: Longman, Inc., p. 97.
60. Dow, T. G. B., Rooney, P. J. y Spence, M. (1975). Does anaemia increase the risks to the fetus caused by smoking in pregnancy? *British Medical Journal* **4**, 253-254.
61. Christophers, S. R. (1924). The mechanism of immunity against malaria in communities living under hyper-endemic conditions. *Indian Journal of Medical Research* **12**, 273-294. Datos reproducidos en Williams, C. B. (1964). *Patterns in the Balance of Nature*. London: Academic Press, p. 243.
62. Datos tomados de *Climatological Data, Ohio y Local Climatological Data, Cleveland, Ohio*; National Oceanic and Atmospheric Administration, U.S. Dept. of Commerce.
63. Estos datos se publicaron en la página 8-A del *Cleveland Plain Dealer*, 6 de febrero de 1997, a partir de información recopilada por la United Network for Organ Sharing. Las variables de tasa de mortalidad y de volumen son promedios de un periodo de cuatro años que empieza en octubre de 1987. Hay 31 hospitales en el grupo de bajo volumen y 76 en el grupo de alto volumen.
64. Erne, P., Bolli, P., Buerigisser, E. y Buehler, F. R. (1984). Correlation of platelet calcium with blood pressure. *New England Journal of Medicine* **310**, 1084-1088. Reimpreso con permiso. Datos originales cortesía de F. R. Buehler. El conjunto de datos original tenía 47 sujetos; hemos omitido 9 pacientes con presiones sanguíneas «en la frontera».

LA PROBABILIDAD Y LA DISTRIBUCIÓN BINOMIAL

Objetivos

En este capítulo estudiaremos las ideas básicas de la probabilidad, entre las que se encuentran

- la definición de probabilidad como «frecuencia límite»;
- el uso de árboles de probabilidad;
- el concepto de variable aleatoria;
- las reglas para obtener las medias y las desviaciones típicas de variables aleatorias;
- el uso de la distribución binomial.

3.1 La probabilidad y las Ciencias de la Vida

La probabilidad, o azar, juega un papel importante en el pensamiento científico sobre los sistemas vivos. Algunos procesos biológicos están afectados directamente por el azar. Un ejemplo familiar es la separación de los cromosomas en la formación de los gametos. Otro ejemplo es la aparición de mutaciones.

Incluso aunque en el propio proceso biológico no intervenga el azar, los resultados de un experimento están siempre algo afectados por dicho azar: fluctuaciones aleatorias en las condiciones experimentales, variaciones aleatorias en la genética de los animales experimentales, etc. A menudo, el azar también interviene directamente en el diseño del experimento. Por ejemplo, diversas variedades de trigo se pueden asignar aleatoriamente a las parcelas de un experimento. (La asignación aleatoria se presentará en el Capítulo 11).

Las conclusiones de un análisis de datos estadístico se establecen a menudo en términos de probabilidad. La probabilidad interviene en el análisis estadístico no solo porque el azar influye en los resultados del experimento, sino también porque los modelos de probabilidad nos permiten cuantificar cómo es de probable o improbable un resultado experimental, dadas ciertas suposiciones de modelado. En este capítulo presentaremos el lenguaje de la probabilidad y desarrollaremos algunas herramientas simples para manejar probabilidades.

3.2 Introducción a la probabilidad

En esta sección presentaremos el lenguaje de la probabilidad y su interpretación.

Conceptos básicos

Una **probabilidad** es una cantidad numérica que expresa la verosimilitud de un suceso. La probabilidad de un suceso E se expresa como

$$\Pr\{E\}$$

La probabilidad $\Pr\{E\}$ es siempre un número entre 0 y 1, ambos inclusive.

Solo tiene significado hablar de una probabilidad $\Pr\{E\}$ en el contexto de una operación aleatoria; es decir, una operación cuyo resultado está determinado al menos parcialmente por el azar. La operación aleatoria se puede definir de forma que *cada vez que dicha operación aleatoria se realiza, el suceso E ocurre o no ocurre*. Los dos ejemplos siguientes ilustran estas ideas.

Ejemplo 3.2.1 Lanzamiento de una moneda

Consideremos la operación aleatoria familiar de lanzar una moneda, y definamos el suceso:

E : salir cara

Cada vez que se lance la moneda, o sale cara o no. Si la moneda es tal que es igualmente probable que salgan caras o cruces, entonces

$$\Pr\{E\} = \frac{1}{2} = 0,5$$

Esta moneda ideal se denomina moneda «justa». Si la moneda no es justa (quizá porque esté ligeramente doblada), entonces $\Pr\{E\}$ puede tener un valor distinto de 0,5, por ejemplo,

$$\Pr\{E\} = 0,6$$

■

Ejemplo 3.2.2 Lanzamiento de una moneda

Consideremos el suceso:

E : tres caras seguidas

La operación aleatoria «lanzar una moneda» *no* es adecuada para este suceso, ya que no podemos averiguar con un solo lanzamiento si E ha ocurrido. Una operación aleatoria que sería adecuada es

Operación aleatoria: Lanzar una moneda tres veces.

Otra operación aleatoria que sería adecuada es

Operación aleatoria: Lanzar una moneda 100 veces

entendiéndose que E ocurre si salen tres caras seguidas en cualquier momento de los 100 lanzamientos. La intuición sugiere que E sería más probable en la segunda definición de la operación aleatoria (100 lanzamientos) que en la primera (3 lanzamientos). Esta intuición es correcta y sirve para subrayar la importancia de la operación aleatoria en la interpretación de una probabilidad. ■

El lenguaje de la probabilidad se puede utilizar para describir los resultados del muestreo aleatorio de una población. La aplicación más sencilla de esta idea es un muestreo de tamaño $n = 1$. Es decir, elegir aleatoriamente un miembro de una población. Se presenta a continuación un ejemplo de esta idea.

Ejemplo 3.2.3 Muestreo de moscas de la fruta

Se mantiene en un laboratorio una población grande de la mosca de la fruta *Drosophila melanogaster*. El 30 % de los individuos son negros debido a una mutación. Mientras que el 70 % restante tienen el color de cuerpo gris normal. Supongamos que se escogen aleatoriamente una mosca de la población. Entonces la probabilidad de que se escoja una mosca negra es 0,3. Más formalmente, si definimos

E : la mosca muestreada es negra

Entonces

$$\Pr\{E\} = 0,3$$

■

El ejemplo anterior ilustra la relación básica entre probabilidad y muestreo aleatorio: *la probabilidad de que un individuo escogido aleatoriamente tenga una cierta característica es igual a la proporción de los miembros de la población con dicha característica*.

Interpretación de la probabilidad como frecuencia

La interpretación de la **probabilidad como frecuencia** proporciona una conexión entre la probabilidad y el mundo real, relacionando la probabilidad de un suceso con una cantidad medible, concretamente, la frecuencia relativa de incidencia a largo plazo del suceso*.

De acuerdo con la interpretación como frecuencia, la probabilidad de un suceso E tiene significado solo en relación con una operación aleatoria que en principio se repite indefinidamente. Cada vez que se repite la operación aleatoria, el suceso E ocurre o no ocurre. *La probabilidad $\Pr\{E\}$ se interpreta como la frecuencia relativa de ocurrencias de E en una serie infinitamente larga de repeticiones de la operación aleatoria.*

Concretamente, supongamos que la operación aleatoria se repite un gran número de veces, y que en cada repetición se anota la incidencia o no incidencia de E . Entonces podemos expresar

$$\Pr\{E\} \leftrightarrow \frac{\# \text{ de veces que ocurre } E}{\# \text{ de veces que se repite la operación aleatoria}}$$

La flecha de la expresión anterior indica «igualdad aproximada a largo plazo». Es decir, si la operación aleatoria se repitiera muchas veces, los dos miembros de la expresión serían aproximadamente iguales. Presentamos a continuación un ejemplo sencillo.

Ejemplo 3.2.4 Lanzamiento de una moneda

Consideremos de nuevo la operación aleatoria de lanzar una moneda, y el suceso

E : salir cara

Si la moneda es justa, entonces

$$\Pr\{E\} = 0,5 \leftrightarrow \frac{\# \text{ de caras}}{\# \text{ de cruces}}$$

La flecha en la expresión anterior indica que, en una larga serie de lanzamientos de una moneda justa, podemos esperar obtener cara aproximadamente el 50 % de las veces. ■

Los dos ejemplos siguientes ilustran la interpretación de frecuencias relativas en sucesos más complejos.

Ejemplo 3.2.5 Lanzamiento de una moneda

Supongamos que se lanza dos veces una moneda justa. Por razones que se explicará más adelante en esta sección, la probabilidad de obtener cara las dos veces es 0,25. Esta probabilidad tiene la siguiente interpretación de frecuencia relativa

Operación aleatoria: lanzar dos veces una moneda

E : en ambos lanzamientos sale cara

$$\Pr\{E\} = 0,25 \leftrightarrow \frac{\# \text{ de veces que sale cara las dos veces}}{\# \text{ de veces que se lanzan las dos monedas}}$$

Ejemplo 3.2.6 Muestreo de moscas de la fruta

En la población *Drosophila* del Ejemplo 3.2.3, el 30 % de las moscas son negras y el 70 % son grises. Supongamos que se escogen aleatoriamente dos moscas de la población. Veremos más adelante en esta sección que la probabilidad de que ambas moscas tengan el mismo color es de 0,58. Esta probabilidad se puede interpretar de la siguiente forma:

Operación aleatoria: elegir una muestra aleatoria de tamaño $n = 2$

E : las dos moscas de la muestra son del mismo color

$$\Pr\{E\} = 0,58 \leftrightarrow \frac{\# \text{ de veces que ambas moscas son del mismo color}}{\# \text{ de veces que se elige una muestra de tamaño } n = 2}$$

* Algunos estadísticos prefieren una perspectiva diferente, concretamente que la probabilidad de un suceso es una cantidad subjetiva que expresa el «grado de confianza» en que un suceso ocurrirá. Los métodos estadísticos basados en esta interpretación «subjetivista» son algo diferentes a los presentados en este libro.

Podemos relacionar esta interpretación con un experimento de muestreo concreto. Supongamos que la población de *Drosophila* está en un contenedor muy grande, y que tenemos algún mecanismo que permite escoger aleatoriamente una mosca del contenedor. Escogemos una mosca aleatoriamente y después otra. Estas dos moscas constituyen la primera muestra de $n = 2$. Tras apuntar sus colores, volvemos a poner las dos moscas en el contenedor, y estamos listos para repetir la operación de muestreo de nuevo otra vez. Este experimento de muestreo sería tedioso de realizar físicamente, pero se puede simular fácilmente utilizando un computador. La Tabla 3.2.1 muestra un registro parcial de los resultados de escoger 10.000 muestras aleatorias de tamaño $n = 2$ de una población simulada de *Drosophila*. Tras cada repetición de la operación aleatoria (es decir, tras cada muestra de $n = 2$), se actualiza la frecuencia relativa del suceso E , como se indica en la columna de la tabla que está más a la derecha.

Tabla 3.2.1 Resultados parciales de muestreo aleatorio de una población de <i>Drosophila</i>				
Número de muestra	Color		¿Ocurrió E ?	Frecuencia relativa de E (acumulativa)
	Primera mosca	Segunda mosca		
1	G	B	No	0,000
2	B	B	Sí	0,500
3	B	G	No	0,333
4	G	B	No	0,250
5	G	G	Sí	0,400
6	G	B	No	0,333
7	B	B	Sí	0,429
8	G	G	Sí	0,500
9	G	B	No	0,444
10	B	B	Sí	0,500
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
20	G	B	No	0,450
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
100	G	B	No	0,540
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1.000	G	G	Sí	0,596
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
10.000	B	B	Sí	0,577

La Figura 3.2.1 muestra la frecuencia relativa actualizada en función del número de muestras. Nótese que, cuando el número de muestras se hace grande, la frecuencia relativa de incidencias de E se acerca a 0,58 (que es $\Pr\{E\}$). En otras palabras, el porcentaje de muestras de color homogéneo en todas las muestras se acerca al 58 % a medida que el número de muestras aumenta. Sin embargo, debe resaltarse que el número *absoluto* de muestras de color homogéneo generalmente no tiende a acercarse al 58 % del número total. Por ejemplo, si comparamos los

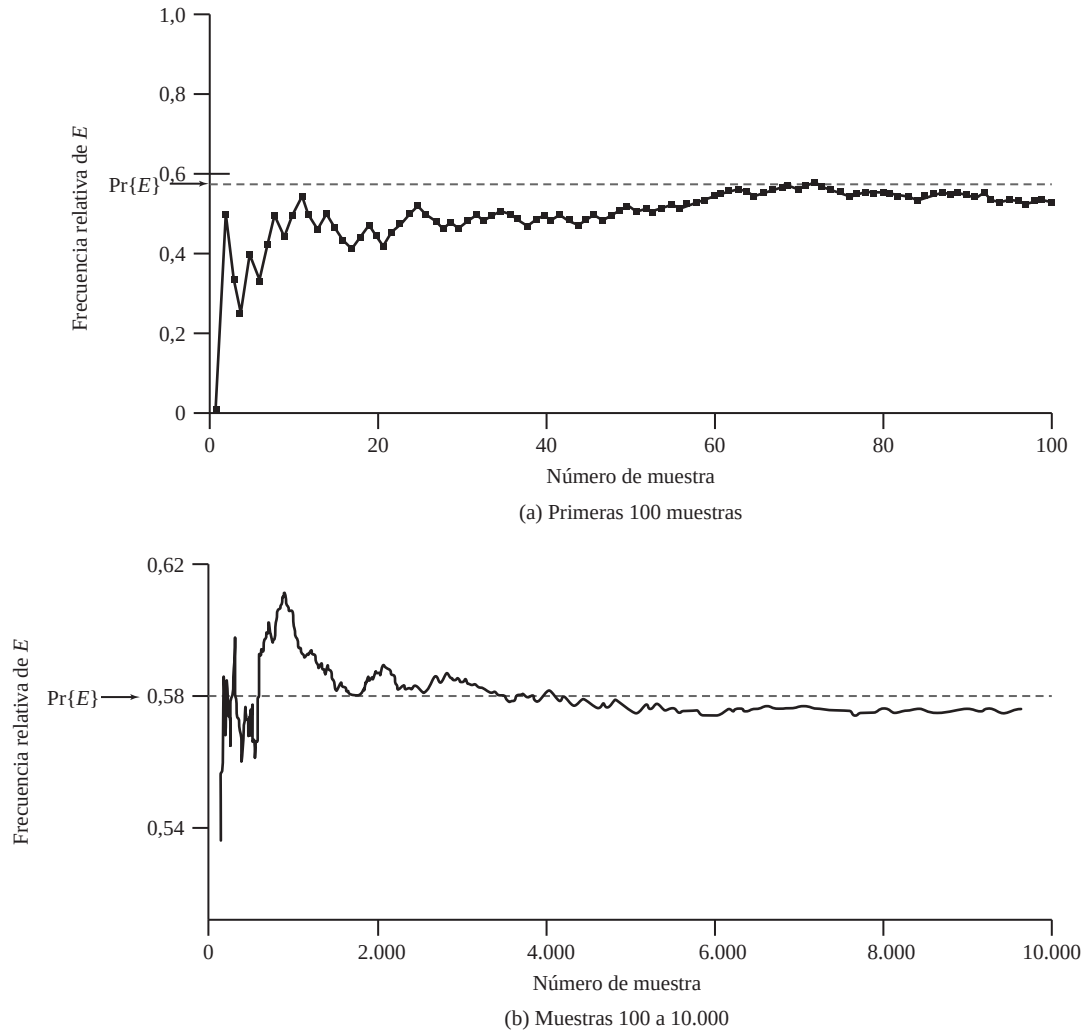


Figura 3.2.1 Resultados del muestreo de la población de mosca de la fruta. Nótese que en (a) y (b) la escala de los ejes es diferente

resultados que se muestran en la Tabla 3.2.1 para las primeras 100 muestras y para las primeras 1.000 muestras, encontramos lo siguiente:

	Desviación del 58 % o del total					
	De color homogéneo					
Primeras 100 muestras:	54	o	54 %	-4	o	-4 %
Primeras 1.000 muestras:	596	o	59,6 %	+16	o	+1,6 %

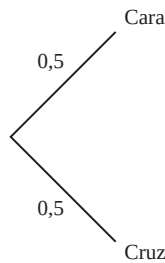
Nótese que la desviación del 58 % es mayor en términos absolutos, pero menor en términos relativos (es decir, en términos de porcentajes), para 1.000 muestras que para 100 muestras. Asimismo, para 10.000 muestras la desviación del 58 % es aún mayor (una desviación de -30), pero la desviación en porcentaje es muy pequeña (30/10.000 es 0,3 %). El déficit de cuatro muestras de color homogéneo en las 100 primeras muestras no se cancela por un correspondiente exceso en muestras posteriores, sino que es reducido al dividir la frecuencia de aparición por un denominador mayor. ■

Árboles de probabilidad

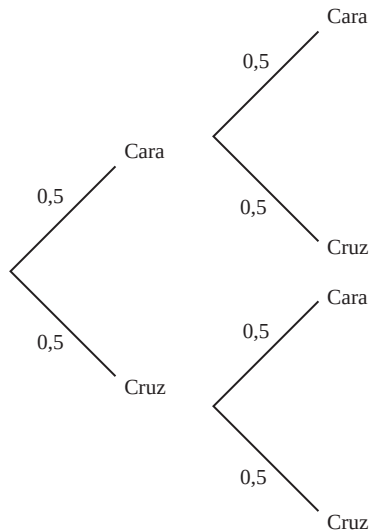
A menudo es útil utilizar un **árbol de probabilidad** para analizar un problema de probabilidades. Un árbol de probabilidad proporciona una forma adecuada de dividir un problema en partes y organizar la información disponible. Los siguientes ejemplos muestran algunas aplicaciones de esta idea.

Ejemplo 3.2.7 Lanzamiento de una moneda

Si se lanza dos veces una moneda justa, entonces la probabilidad de obtener cara es 0,5 en cada lanzamiento. La primera parte de un árbol de probabilidad en este escenario muestra que hay dos posibles resultados del primer lanzamiento y que cada uno de ellos tiene una probabilidad de 0,5.



Después el árbol muestra que, para cada resultado del primer lanzamiento, el segundo lanzamiento también puede ser cara o cruz, de nuevo con probabilidad 0,5 en cada caso.



Para obtener la probabilidad de obtener cara en ambos lanzamientos, consideraremos el camino por el árbol que produce este suceso. Multiplicaremos entre sí las probabilidades que encontramos a lo largo de ese camino. La Figura 3.2.2 resume este ejemplo y muestra que

$$\Pr\{\text{cara en ambos lanzamientos}\} = 0,5 \times 0,5 = 0,25$$



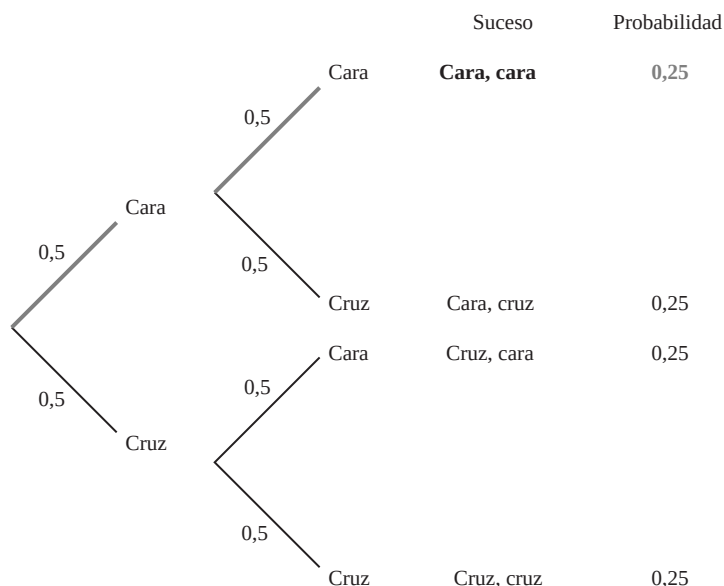


Figura 3.2.2 Árbol de probabilidad del lanzamiento de dos monedas

Combinación de probabilidades

Si un suceso puede ocurrir en más de una forma, la interpretación de frecuencia relativa de la probabilidad puede ser una guía para combinar apropiadamente las probabilidades de los subsucesos. El siguiente ejemplo ilustra esta idea.

Ejemplo 3.2.8 Muestreo de mosca de la fruta

En los Ejemplos 3.2.3 y 3.2.6 de la población de *Drosophila*, el 30 % de las moscas son negras y el 70 % son grises. Supongamos que se escogen aleatoriamente dos moscas de la población. Supongamos además que deseamos obtener la probabilidad de que ambas moscas sean del mismo color. El árbol de probabilidad que se muestra en la Figura 3.2.3 indica los cuatro posibles resultados de tomar una muestra de dos moscas. Observando el árbol, podemos ver que la probabilidad de obtener dos moscas negras es $0,3 \times 0,3 = 0,09$. Asimismo, la probabilidad de obtener dos moscas grises es $0,7 \times 0,7 = 0,49$.

Para obtener la probabilidad del suceso

E : ambas moscas en la muestra son del mismo color

sumamos la probabilidad de negro, negro a la probabilidad de gris, gris y obtenemos $0,09 + 0,49 = 0,58$. ■

En el experimento de lanzamiento de moneda del Ejemplo 3.2.7, la segunda parte del árbol de probabilidad tenía la misma estructura que la primera parte (una probabilidad de 0,5 de obtener cara y una probabilidad de 0,5 de obtener cruz), ya que el resultado del primer lanzamiento no afecta a la probabilidad de obtener cara en el segundo lanzamiento. Asimismo, en el Ejemplo 3.2.8, la probabilidad de que la segunda mosca sea negra era de 0,3, independientemente del color de la primera mosca, porque la población se suponía muy grande, de manera que extraer una mosca de la población no afecta a la proporción de moscas que son negras. Sin embargo, en algunas situaciones es necesario tratar la segunda parte del árbol de probabilidad de forma diferente que la primera parte.

Ejemplo 3.2.9 Óxido nítrico

El fallo respiratorio por hipoxia es una situación seria que afecta a algunos recién nacidos. Si un recién nacido está en esta situación, a menudo es necesario utilizar oxigenación mediante membrana extracorpórea (ECMO) para

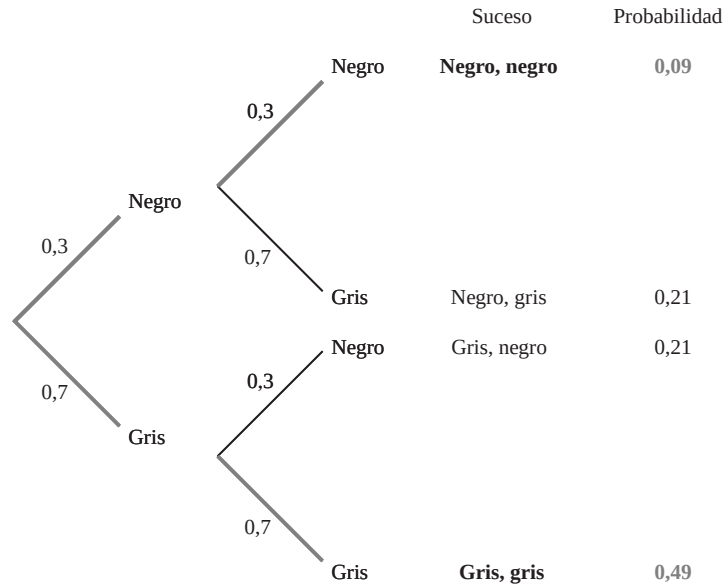


Figura 3.2.3 Árbol de probabilidad del muestreo de dos moscas

salvar la vida del niño. Sin embargo, la ECMO es un procedimiento invasivo en el que se inserta un tubo en una vena o arteria cerca del corazón, por lo que los médicos intentan evitar su necesidad. Un tratamiento del fallo respiratorio hipóxico es hacer que el recién nacido inhale óxido nítrico. Para probar la efectividad este tratamiento, se asignaron aleatoriamente recién nacidos que sufrían fallo respiratorio hipóxico a un grupo que recibía óxido nítrico o a un grupo de control¹. En el grupo de tratamiento, el 45,6 % de los recién nacidos tuvieron un resultado negativo, lo que significa que o bien necesitaron ECMO o fallecieron. En el grupo de control, el 63,6 % de los recién nacidos tuvieron un resultado negativo. La Figura 3.2.4 muestra el árbol de probabilidad de este experimento.

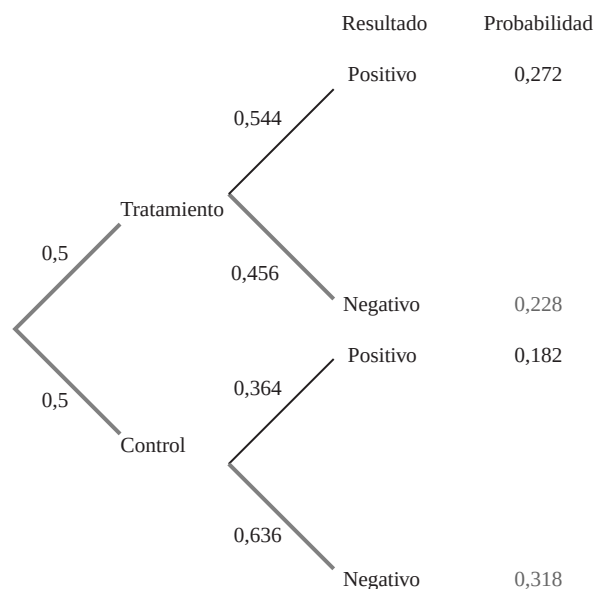


Figura 3.2.4 Árbol de probabilidad del ejemplo del óxido nítrico

Si seleccionamos de forma aleatoria un recién nacido de este grupo, hay una probabilidad de 0,5 de que el recién nacido esté en el grupo de tratamiento y, si es así, una probabilidad de 0,456 de obtener un resultado negativo. Asimismo, hay una probabilidad de 0,5 de que el recién nacido esté en el grupo de control y, si es así, una probabilidad de 0,636 de obtener un resultado negativo. Por tanto, la probabilidad de un resultado negativo es

$$0,5 \times 0,456 + 0,5 \times 0,636 = 0,228 + 0,318 = 0,546$$

Ejemplo 3.2.10 Prueba médica

Supongamos que se realiza una prueba médica en una persona para intentar determinar si dicha persona tiene o no tiene una enfermedad en particular. Si la prueba indica que la enfermedad está presente, se dice que dicha persona tiene un «resultado positivo». Si la prueba indica que la enfermedad no está presente, se dice que dicha persona tiene un «resultado negativo». Sin embargo, hay dos tipos de errores que se pueden cometer. Es posible que la prueba indique que la enfermedad está presente, pero que la persona realmente no tenga la enfermedad; esto se conoce como falso positivo. También es posible que la persona tenga la enfermedad, pero que la prueba no la detecte. Esto se conoce como falso negativo.

Supongamos que una prueba concreta tiene un 95 % de probabilidades de detectar la enfermedad si la persona la tiene (esto se denomina sensibilidad de la prueba), y el 90 % de probabilidades de indicar correctamente que la enfermedad está ausente si la persona realmente no tiene dicha enfermedad (esto se denomina especificidad de la prueba). Supongamos que el 8 % de la población tiene la enfermedad. ¿Cuál es la probabilidad de que una persona escogida aleatoriamente dé un resultado positivo?

La Figura 3.2.5 muestra un árbol de probabilidad de esta situación. La primera bifurcación del árbol muestra la división entre aquellos que tienen la enfermedad y aquellos que no la tienen. Si alguien tiene la enfermedad, entonces utilizamos el 95 % de la probabilidad de que la persona dé un resultado positivo. Si la persona no tiene la enfermedad, en todos utilizamos el 0,10 de la probabilidad de que la persona dé un resultado positivo. Por tanto, la probabilidad de que una persona escogida aleatoriamente dé un resultado positivo es

$$0,08 \times 0,95 + 0,92 \times 0,10 = 0,076 + 0,092 = 0,168$$

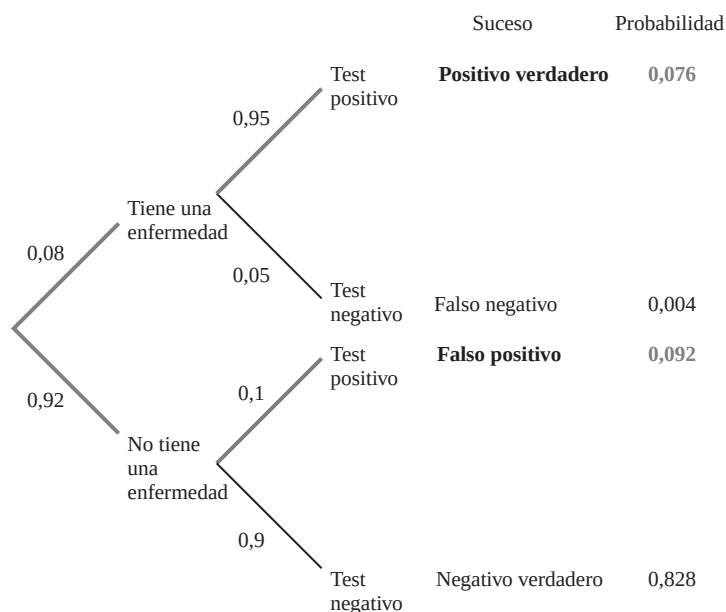


Figura 3.2.5 Árbol de probabilidad del ejemplo del tratamiento médico

Ejemplo 3.2.11 Falsos positivos

Consideremos el escenario de la prueba médica del Ejemplo 3.2.10. Si el resultado de la prueba sobre alguien es positivo, ¿cuál es la probabilidad de que la persona realmente tenga la enfermedad? En el Ejemplo 3.2.10 obtuvimos que 0,168 (el 16,8 %) de la población tendría un resultado de la prueba positivo, de forma que si se realizara la prueba a 1.000 personas, se podría esperar que 168 tuvieran un resultado positivo. La probabilidad de un positivo verdadero es de 0,076, por lo que se podría esperar 76 «positivos verdaderos» en las 1.000 personas sobre las que se realizara la prueba. Por tanto, se pueden esperar 76 verdaderos positivos de 168 positivos en total, es decir que la probabilidad de que alguien realmente tenga la enfermedad, dado que el resultado de la prueba sobre dicha persona

es positivo, es $\frac{76}{168} = \frac{0,076}{0,168} \approx 0,452$. Esta probabilidad es un tanto pequeña respecto a la que la mayoría de la gente esperaría obtener, dado que la sensibilidad y la especificidad del texto son de 0,95 y 0,90. ■

Ejercicios 3.2.1-3.2.7

3.2.1 En una cierta población de cabezas gordas de agua dulce, *Cottus rotheus*, la distribución del número de vértebras en la cola se muestra en la tabla².

Número de vértebras	Porcentaje de peces
20	3
21	51
22	40
23	6
Total	100

Calcule la probabilidad de que el número de vértebras en la cola de un pez escogido aleatoriamente de la población

- (a) sea igual a 21;
- (b) sea menor o igual que 22;
- (c) sea mayor que 21;
- (d) no sea mayor que 21.

3.2.2 En una cierta universidad, el 55 % de los estudiantes son mujeres. Supongamos que tomamos una muestra de dos estudiantes. Utilice un árbol de probabilidad para obtener la probabilidad

- (a) de que los dos estudiantes elegidos sean mujeres;
- (b) de que al menos uno de los dos estudiantes sea una mujer.

3.2.3 Suponga que una enfermedad se hereda de manera ligada al sexo, de forma que la descendencia masculina tiene un 50 % de probabilidades de heredar la enfermedad, pero la descendencia femenina no tiene ninguna probabilidad de heredar la enfermedad. Suponga además que el 51,3 % de los niños son varones. ¿Cuál es la probabilidad de que un niño escogido aleatoriamente esté afectado por la enfermedad?

3.2.4 Suponga que un estudiante que está a punto de realizar un test de múltiples respuestas solo ha aprendido el 40 % de la materia del examen. Por tanto, hay una probabilidad del 40 % de que sepa la respuesta a una pregunta. Sin embargo, incluso aunque no sepa la respuesta a una pregunta, tiene todavía una probabilidad del 20 % de obtener la respuesta correcta adivinándola. Si se escoge aleatoriamente una pregunta del examen, ¿cuál es la probabilidad de que su respuesta sea correcta?

3.2.5 Si una mujer realiza una prueba de embarazo, puede obtener un resultado positivo, lo que significa que la prueba indica que está embarazada o un resultado negativo, lo que significa que la prueba dice que no está embarazada. Supongamos que si una mujer realmente está embarazada, hay un 98 % de probabilidades de que la prueba dé un resultado positivo. Supongamos también que si una mujer realmente *no* está embarazada, hay un 99 % de probabilidades de que la prueba dé un resultado negativo.

- (a) Suponga que 1.000 mujeres realizan la prueba de embarazo y que 100 de ellas realmente están embarazadas. ¿Cuál es la probabilidad de que una mujer escogida aleatoriamente de este grupo dé un resultado positivo?
- (b) Suponga que 1.000 mujeres realizan la prueba de embarazo y que 50 realmente están embarazadas. ¿Cuál es la probabilidad de que una mujer elegida aleatoriamente de este grupo dé un resultado positivo?

3.2.6

- (a) Considere el experimento del Ejercicio 3.2.5, apartado (a). Suponga que el resultado de la mujer es positivo. ¿Cuál es la probabilidad de que realmente esté embarazada?
- (b) Considere el experimento del Ejercicio 3.2.5, apartado (b). Suponga que el resultado de la mujer es positivo. ¿Cuál es la probabilidad de que realmente esté embarazada?

3.2.7 Suponga que una prueba médica tiene un 92 % de probabilidades de detectar una enfermedad si la persona la tiene (es decir, una sensibilidad del 92 %) y un 94 % de probabilidades de indicar correctamente que la enfermedad está ausente si la persona realmente no tiene dicha enfermedad (es decir, una especificidad del 94 %). Suponga que el 10 % de la población tiene la enfermedad.

- ¿Cuál es la probabilidad de que una persona escogida aleatoriamente dé un resultado positivo?
- Suponga que una persona escogida aleatoriamente da un resultado positivo. ¿Cuál es la probabilidad de que esta persona realmente tenga la enfermedad?

3.3 Las reglas de la probabilidad (opcional)

Hemos definido la probabilidad de un suceso, $\Pr\{E\}$, como la frecuencia relativa a largo plazo con la que dicho suceso ocurre. En esta sección consideraremos brevemente algunas reglas que ayudan a determinar probabilidades. Comenzaremos con tres reglas básicas.

Reglas básicas

Regla (1). La probabilidad de un suceso E está siempre entre 0 y 1. Es decir, $0 \leq \Pr\{E\} \leq 1$.

Regla (2). La suma de las probabilidades de todos los posibles sucesos es igual a 1. Es decir, si el conjunto de posibles sucesos es $E_1, E_2, \dots, E_k$, entonces $\sum_{i=1}^k \Pr\{E_i\} = 1$.

Regla (3). La probabilidad de que un suceso E no ocurra, que se indica como E^C , es uno menos la probabilidad de que el suceso ocurra. Es decir, $\Pr\{E^C\} = 1 - \Pr\{E\}$. (E^C se denomina suceso complementario o contrario de E).

Ilustraremos estas reglas con un ejemplo.

Ejemplo 3.3.1 Grupo sanguíneo

En Estados Unidos, el 44 % de la población tiene grupo sanguíneo O, el 42 % tiene grupo A, el 10 % tiene grupo B y el 4 % tiene grupo AB³. Consideremos que se elige a una persona aleatoriamente y se determina su grupo sanguíneo. La probabilidad del grupo sanguíneo dado corresponderá al porcentaje de la población.

- La probabilidad de que la persona tenga grupo sanguíneo O = $\Pr\{O\} = 0,44$.
- $\Pr\{O\} + \Pr\{A\} + \Pr\{B\} + \Pr\{AB\} = 0,44 + 0,42 + 0,10 + 0,04 = 1$.
- La probabilidad de que una persona *no* tenga grupo sanguíneo O = $\Pr\{O^C\} = 1 - 0,44 = 0,56$. Este resultado se podría haber obtenido también sumando las probabilidades de los otros grupos sanguíneos: $\Pr\{O^C\} = \Pr\{A\} + \Pr\{B\} + \Pr\{AB\} = 0,42 + 0,10 + 0,04 = 0,56$. ■

Frecuentemente desearíamos considerar dos o más sucesos a la vez. Para ello encontraremos útil el uso de alguna terminología. Diremos que dos sucesos son *disjuntos** si no pueden ocurrir simultáneamente. La Figura 3.3.1 es un *diagrama de Venn* que muestra el *espacio muestral* S de todos los posibles resultados como un rectángulo con dos sucesos disjuntos que se muestran como regiones que no se solapan.

La *unión* de dos sucesos es el suceso que ocurre cuando ocurre uno de los dos o ambos. La *intersección* de dos sucesos es el suceso que ocurre cuando ambos sucesos ocurren. La Figura 3.3.2 es un diagrama de Venn que muestra la unión de dos sucesos como el área total sombreada, y la intersección de los sucesos como la región de solapamiento en el centro.

Si dos sucesos son disjuntos, entonces la probabilidad de su unión es la suma de sus probabilidades individuales. Si los sucesos no son disjuntos, entonces para obtener la probabilidad de su unión tomaremos la suma de sus probabilidades individuales y restaremos la probabilidad de su intersección (la parte que se ha «contado dos veces»).

* Otro término para los sucesos disjuntos es sucesos «mutuamente excluyentes».

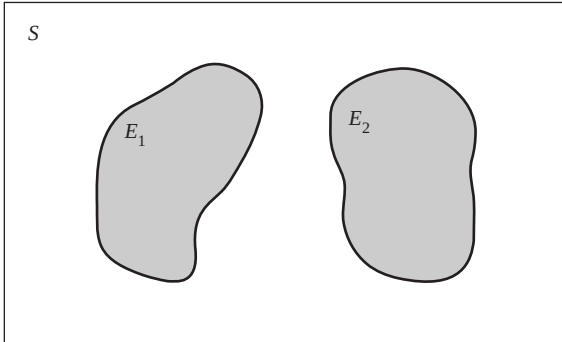


Figura 3.3.1 Diagrama de Venn que muestra dos sucesos disjuntos

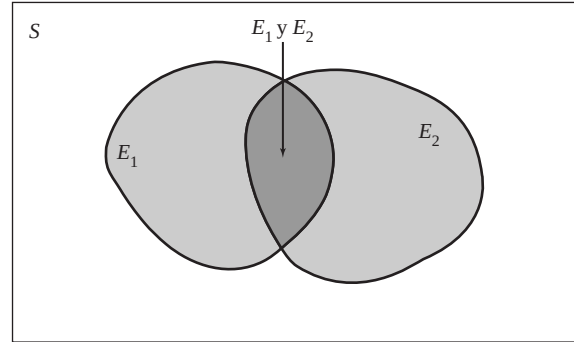


Figura 3.3.2 Diagrama de Venn que muestra la unión (área total sombreada) y la intersección (área central) de dos sucesos

Reglas de la suma

Regla (4). Si dos sucesos E_1 y E_2 son disjuntos, entonces
 $\Pr\{E_1 \text{ o } E_2\} = \Pr\{E_1\} + \Pr\{E_2\}$.

Regla (5). Para dos sucesos cualesquiera E_1 y E_2 , entonces
 $\Pr\{E_1 \text{ o } E_2\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1 \text{ y } E_2\}$.

Ilustraremos estas reglas con un ejemplo.

Ejemplo 3.3.2 Color de cabello y color de ojos

La Tabla 3.3.1 muestra la relación entre el color del pelo y el color de los ojos de un grupo de 1.770 hombres alemanes⁴.

		Color de cabello			Total
		Castaño	Negro	Rojo	
Color de ojos	Castaño	400	300	20	720
	Azules	800	200	50	1.050
	Total	1.200	500	70	1.770

- Como los sucesos «pelo negro» y «pelo rojo» son disjuntos, si se elige alguien de este grupo de forma aleatoria, entonces $\Pr\{\text{pelo negro o pelo rojo}\} = \Pr\{\text{pelo negro}\} + \Pr\{\text{pelo rojo}\} = 500/1.770 + 70/1.770 = 570/1.770$.
- Si se escoge aleatoriamente a una persona de este grupo, entonces $\Pr\{\text{pelo negro}\} = 500/1.770$.
- Si se escoge aleatoriamente a una persona de este grupo, entonces $\Pr\{\text{ojos azules}\} = 1.050/1.770$.
- Los sucesos «pelo negro» y «ojos azules» no son disjuntos, ya que hay 200 hombres que tienen pelo negro y ojos azules. Por tanto, $\Pr\{\text{pelo negro u ojos azules}\} = \Pr\{\text{pelo negro}\} + \Pr\{\text{ojos azules}\} - \Pr\{\text{pelo negro y ojos azules}\} = 500/1.770 + 1.050/1.770 - 200/1.770 = 1.350/1.770$. ■

Se dice que dos sucesos son *independientes* si el conocimiento de que ha ocurrido uno de ellos no cambia la probabilidad de que el otro ocurra. Por ejemplo, si se lanza una moneda dos veces, el resultado del segundo lanzamiento es independiente del resultado del primer lanzamiento, ya que saber que el primer resultado fue cara o cruz no cambia la probabilidad de obtener cara en el segundo lanzamiento.

Los sucesos que no son independientes se denominan *dependientes*. Cuando los sucesos son dependientes, es necesario considerar la *probabilidad condicional* de un suceso, dado que el otro suceso ha ocurrido. Utilizamos la notación

$$\Pr\{E_2|E_1\}$$

para representar la probabilidad de que suceda E_2 , dado que ha ocurrido E_1 .

Ejemplo 3.3.3 Color de cabello y color de ojos

Consideremos que se escoge aleatoriamente un hombre del grupo que muestra la Tabla 3.3.1. Globalmente, la probabilidad de ojos azules es $1.050/1.770$, o aproximadamente el 59,3 %. Sin embargo, si el hombre tiene pelo negro, entonces la probabilidad condicional de ojos azules es solo $200/500$, o el 40 %. Es decir, $\Pr\{\text{pelo negro}|\text{ojos azules}\} = 0,40$. Como la probabilidad de ojos azules depende del color del cabello, por lo que los sucesos «pelo negro» y «ojos azules» son dependientes. ■

Refirámonos de nuevo a la Figura 3.3.2, que muestra la intersección de dos regiones (de E_1 y E_2). Si sabemos que el suceso E_2 ha ocurrido, entonces podemos restringir nuestra atención a la región de E_1 del diagrama de Venn. Si ahora deseamos calcular la probabilidad de que suceda E_2 , necesitamos considerar la intersección de E_1 y E_2 relativa a la región entera de E_1 . En el caso del Ejemplo 3.3.3, esto corresponde a saber que un hombre escogido aleatoriamente tiene pelo negro, de forma que restringimos nuestra atención a los 500 hombres (del total de 1.770 del grupo) con pelo negro. De esos hombres, 200 tienen ojos azules. Los 200 están en la intersección de «pelo negro» y «ojos azules». La fracción $200/500$ es la probabilidad condicional de tener ojos azules, dado que el hombre tiene pelo negro.

Esto conduce a la siguiente definición formal de probabilidad condicional de E_2 dado E_1 .

Definición La probabilidad condicional de E_2 , dado E_1 es

$$\Pr\{E_2|E_1\} = \frac{\Pr\{E_1 \text{ y } E_2\}}{\Pr\{E_1\}}$$

Supuesto que $\Pr\{E_1\} > 0$.

Ejemplo 3.3.4 Color de cabello y color de ojos

Consideremos que se escoge un hombre aleatoriamente del grupo que se muestra en la Tabla 3.3.1. La probabilidad de que el hombre tenga ojos azules dado que tiene pelo negro es

$$\begin{aligned}\Pr\{\text{ojos azules}|\text{pelo negro}\} &= \Pr\{\text{pelo negro y ojos azules}\}/\Pr\{\text{pelo negro}\} \\ &= \frac{200/1.770}{500/1.770} = \frac{200}{500} = 0,40\end{aligned}$$

En la Sección 3.2 utilizamos árboles de probabilidad para estudiar sucesos compuestos. Al hacerlo así, utilizamos implícitamente en las reglas de la multiplicación que haremos explícitas a continuación.

Reglas de la multiplicación

Regla (6). Si dos sucesos E_1 y E_2 son independientes, entonces

$$\Pr\{E_1 \text{ y } E_2\} = \Pr\{E_1\} \times \Pr\{E_2\}.$$

Regla (7). Para dos sucesos cualesquiera E_1 y E_2 , $\Pr\{E_1 \text{ y } E_2\} = \Pr\{E_1\} \times \Pr\{E_2|E_1\}$.

Ejemplo 3.3.5 Lanzamiento de moneda

Si se lanza dos veces una moneda justa, los dos lanzamientos son independientes entre sí. Por tanto, la probabilidad de obtener cara en ambos lanzamientos es

$$\begin{aligned}\Pr\{\text{dos veces cara}\} &= \Pr\{\text{cara en el primer lanzamiento}\} \times \Pr\{\text{cara en el segundo lanzamiento}\} \\ &= 0,5 \times 0,5 = 0,25\end{aligned}$$

Ejemplo 3.3.6 Grupo sanguíneo

En el Ejemplo 3.3.1 se dijo que el 44 % de la población de Estados Unidos tiene grupo sanguíneo O. También es cierto que el 15 % de la población es Rh negativo y que esto es independiente del grupo sanguíneo. Por tanto, si se escoge aleatoriamente a una persona, la probabilidad de que dicha persona tenga grupo sanguíneo O y Rh negativo es

$$\begin{aligned}\Pr\{\text{grupo O y Rh negativo}\} &= \Pr\{\text{grupo O}\} \times \Pr\{\text{Rh negativo}\} \\ &= 0,44 \times 0,15 = 0,066\end{aligned}$$

Ejemplo 3.3.7 Color de cabello y color de ojos

Consideremos que se escoge aleatoriamente a un hombre del grupo que se muestra en la Tabla 3.3.1. ¿Cuál es la probabilidad de que el hombre tenga pelo rojo y ojos marrones? El color del pelo y color de los ojos son dependientes, de forma que para obtener esta probabilidad hay que utilizar la probabilidad condicional. La probabilidad de que el hombre tenga pelo rojo es 70/1.770. Dado que el hombre tiene pelo rojo, la probabilidad condicional de ojos marrones es 20/70. Por tanto,

$$\begin{aligned}\Pr\{\text{pelo rojo y ojos marrones}\} &= \Pr\{\text{pelo rojo}\} \Pr\{\text{ojos marrones}|\text{pelo rojo}\} \\ &= 70/1.770 \times 20/70 = 20/1.770\end{aligned}$$

Algunas veces un problema de probabilidades se puede dividir en dos «partes» condicionales que se resuelven separadamente y se combinan las respuestas.

Regla de la probabilidad total

Regla (8). Para dos sucesos cualesquiera E_1 y E_2 ,

$$\Pr\{E_1\} = \Pr\{E_2\} \times \Pr\{E_1|E_2\} + \Pr\{E_2^C\} \times \Pr\{E_1|E_2^C\}.$$

Ejemplo 3.3.8 Tamaño de la mano

Consideremos que se escoge aleatoriamente a una persona de una población con un 60 % de mujeres y un 40 % de hombres. Supongamos que para una mujer la probabilidad de tener un tamaño de mano menor que 100 cm² es 0,31⁵. Supongamos que para un hombre la probabilidad de tener un tamaño de mano menor que 100 cm² es 0,08. ¿Cuál es la probabilidad de que una persona elegida aleatoriamente tenga un tamaño de mano menor que 100 cm²?

Si la persona es una mujer, entonces la probabilidad de tamaño de mano «pequeño» es 0,31, y si la persona es un hombre, entonces la probabilidad de tamaño de mano «pequeño» es 0,08.

Entonces,

$$\begin{aligned}\Pr\{\text{tamaño de mano} < 100\} &= \Pr\{\text{mujer}\} \times \Pr\{\text{tamaño de mano} < 100|\text{mujer}\} \\ &\quad + \Pr\{\text{hombre}\} \times \Pr\{\text{tamaño de mano} < 100|\text{hombre}\} \\ &= 0,6 \times 0,31 + 0,4 \times 0,08 \\ &= 0,186 + 0,032 \\ &= 0,218\end{aligned}$$

Ejercicios 3.3.1-3.3.5

3.3.1 En un estudio de la relación entre riesgos para la salud e ingresos, se realizó una serie de preguntas a un gran grupo de gente que vivía en Massachusetts⁶. Algunos de los resultados se muestran en la tabla siguiente

	Ingresos			Total
	Bajos	Medios	Altos	
Fuman	634	332	247	1.213
No fuman	1.846	1.622	1.868	5.336
Total	2.480	1.954	2.115	6.549

- ¿Cuál es la probabilidad de que una persona de este estudio fume?
- ¿Cuál es la probabilidad condicional de que una persona de este estudio fume, sabiendo que dicha persona tiene ingresos altos?
- ¿Es ser un fumador independiente de tener ingresos altos? ¿Por qué o por qué no?

3.3.2 Considere la tabla de datos del Ejercicio 3.3.1.

- ¿Cuál es la probabilidad de que una persona de este estudio sea del grupo de ingresos bajo y fume?
- ¿Cuál es la probabilidad de que una persona de este estudio no sea del grupo de ingresos bajos?
- ¿Cuál es la probabilidad de que una persona de este estudio sea del grupo de ingresos medios?
- ¿Cuál es la probabilidad de que una persona de este estudio sea del grupo de ingresos bajos o del grupo de ingresos medios?

3.3.3 La siguiente tabla de datos está tomada del estudio que se indica en el Ejercicio 3.3.1. En dicha tabla «estresado» significa que la persona ha indicado que la mayor parte de los

días son extremadamente estresantes o bastante estresantes. «No estresado» significa que la persona indica que la mayor parte de los días son poco estresantes, no muy estresantes o en absoluto estresantes.

	Ingresos			Total
	Bajos	Medios	Altos	
Estresados	526	274	216	1.016
No estresados	1.954	1.680	1.899	5.533
Total	2.480	1.954	2.115	6.549

- ¿Cuál es la probabilidad de que una persona de este estudio esté estresada?
- Dado que una persona de este estudio es del grupo de ingresos altos, ¿cuál es la probabilidad de que dicha persona esté estresada?
- Compare sus respuestas a los apartados (a) y (b). ¿Es estar estresado independiente de tener ingresos altos? ¿Por qué o por qué no?

3.3.4 Considere la tabla de datos del Ejercicio 3.3.3.

- ¿Cuál es la probabilidad de que alguien de este estudio tenga ingresos bajos?
- ¿Cuál es la probabilidad de que alguien de este estudio o bien esté estresado o tenga ingresos bajos (o ambas cosas)?
- ¿Cuál es la probabilidad de que alguien en este estudio esté estresado y tenga ingresos bajos?

3.3.5 Suponga que una cierta población de parejas casadas el 30 % de los maridos fuman, el 20 % de las esposas fuman y en el 8 % de las parejas tanto el marido como la esposa fuman. ¿Es la condición de fumador o no fumador del marido independiente de la de la esposa? ¿Por qué o por qué no?

3.4 Curvas de densidad

Los ejemplos presentados en la Sección 3.2 tratan con probabilidades de variables discretas. En esta sección consideraremos la probabilidad cuando la variable es continua.

Histogramas de frecuencias relativas y curvas de densidad

En el Capítulo 2 presentamos el uso del histograma para representar una distribución de frecuencias de una variable. Un histograma de frecuencias relativas es un histograma en el que se indica la proporción (es decir, la frecuencia relativa) de las observaciones en cada categoría, en vez del total de las observaciones en la categoría. Podemos ver el histograma de frecuencias relativas como una aproximación a la verdadera distribución subyacente de la población de la que provienen los datos.

A menudo es deseable, especialmente cuando la variable observada es continua, representar una distribución de frecuencias de una población mediante una curva suave. Podemos ver esta curva como una idealización de un histograma de frecuencias relativas con clases muy estrechas. El siguiente ejemplo ilustra esta idea.

Ejemplo 3.4.1 Glucosa en sangre

Una prueba de tolerancia a la glucosa puede ser útil en el diagnóstico de la diabetes. El nivel de glucosa en sangre se mide una hora después de que el sujeto ha bebido 50 mg de glucosa disuelta en agua. La Figura 3.4.1 muestra la distribución de respuestas a esta prueba de una cierta población de mujeres⁷. La distribución se representa mediante histogramas con anchos de clase iguales a (a) 10, (b) 5 y (c) una curva suave. ■

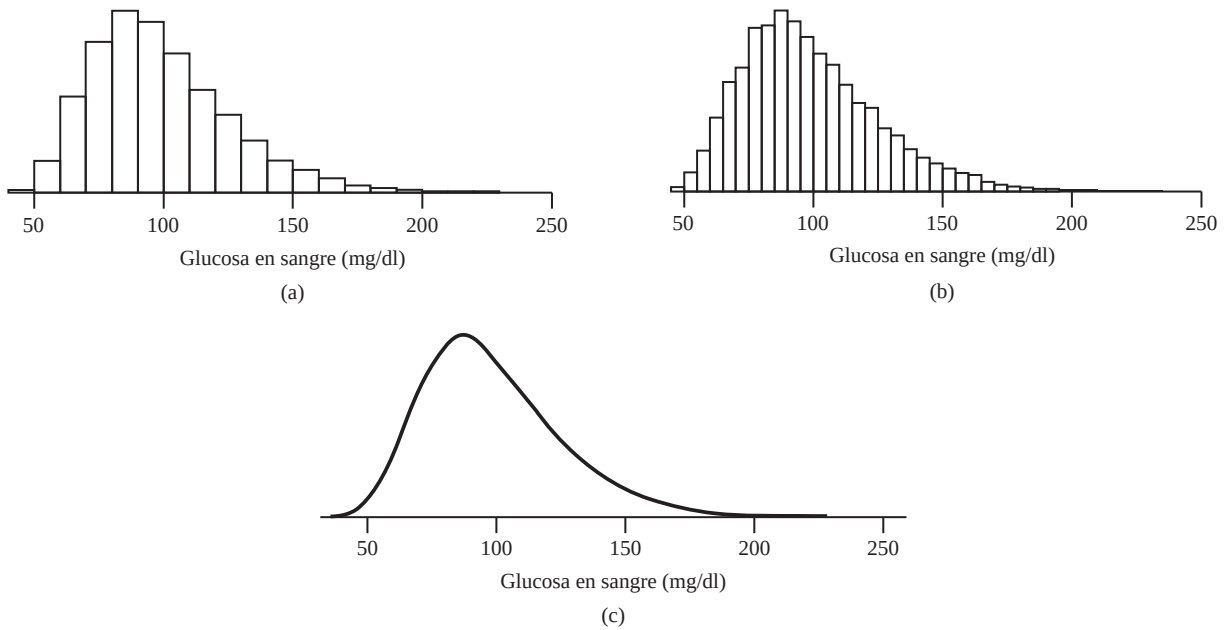


Figura 3.4.1 Diferentes representaciones de la distribución de los niveles de glucosa en sangre en una población de mujeres

Una curva suave que representa una distribución de frecuencias se denomina **curva de densidad**. Las coordenadas verticales de una curva de densidad se representan en una escala denominada **escala de densidades**. Cuando se utiliza la escala de densidades, la frecuencia relativa se representa como áreas bajo la curva. Formalmente, la relación es la siguiente:

Interpretación de la densidad

Dados dos números cualesquiera a y b ,

$\text{Área bajo la curva de densidad entre } a \text{ y } b = \text{Proporción de valores de } Y \text{ entre } a \text{ y } b$

Esta relación se muestra en la Figura 3.4.2 para una distribución arbitraria.

Debido a la forma en la que se interpreta la curva de densidad, dicha curva está completamente por encima (o coincide con) el eje x y el área bajo la curva completa debe ser igual a 1, como se muestra en la Figura 3.4.3.

La interpretación de las curvas de densidad en términos de áreas se ilustra de forma concreta en el siguiente ejemplo.

Ejemplo 3.4.2 Glucosa de sangre

La Figura 3.4.4 muestra la curva de densidad de la distribución de glucosa en sangre del Ejemplo 3.4.1, con la escala vertical mostrada explícitamente. El área sombreada es igual a 0,42, lo que indica que aproximadamente el 42 % de los niveles de glucosa están entre 100 mg/dl y 150 mg/dl. El área bajo la curva de densidad a la izquierda de 100 mg/dl es igual a 0,50, lo que indica que la mediana de la población de niveles de glucosa es de 100 mg/dl. El área bajo la curva completa es uno. ■

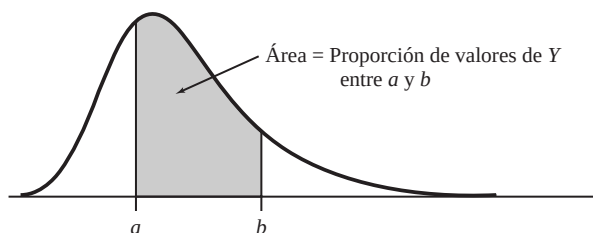


Figura 3.4.2 Interpretación del área bajo una curva de densidad

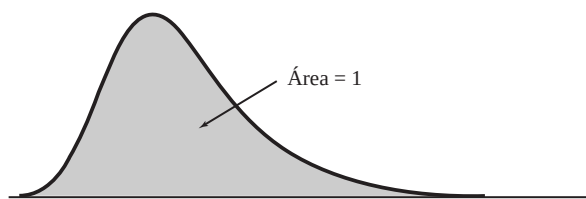


Figura 3.4.3 El área bajo la curva de densidad completa debe ser uno

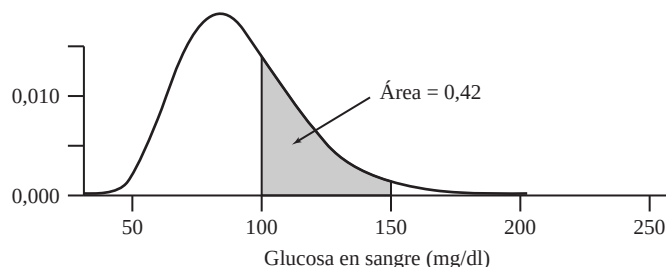


Figura 3.4.4 Interpretación de un área bajo la curva de densidad en el ejemplo de glucosa en sangre

La paradoja del continuo La interpretación como área de una curva de densidad tiene un elemento paradójico: si nos preguntamos por la frecuencia relativa de un valor concreto de Y , la respuesta es cero. Por ejemplo, supongamos que deseamos determinar a partir de la Figura 3.4.4 la frecuencia relativa de los niveles de glucosa en sangre *iguales* a 150. La interpretación del área da una respuesta de cero. Esto parece no tener sentido (¿cómo puede tener cada valor de Y una frecuencia relativa de cero?). Consideremos la cuestión más de cerca. Si el nivel de glucosa en sangre se mide con una precisión de mg/dl, entonces realmente nos estamos preguntando por la frecuencia relativa de los niveles de glucosa entre 149,5 y 150,5 mg/dl, y el área correspondiente no es cero. Por otra parte, si pensamos en los niveles de glucosa como una variable continua *idealizada*, entonces la frecuencia relativa de cualquier valor particular (tal como 150) es cero. La verdad es que es una situación paradójica. Es similar al hecho paradójico de que una línea recta idealizada puede tener una longitud de 1 centímetro, y aun así cada uno de los puntos idealizados de los que se compone la línea tiene una longitud igual a cero. En la práctica, la paradoja del continuo no causa ningún problema; simplemente no consideraremos la frecuencia relativa de un valor concreto de Y (de la misma forma que no se considera la longitud de un único punto).

Probabilidades y curvas de densidad

Si una variable tiene una distribución continua, entonces calcularemos probabilidades utilizando la curva de densidad de dicha variable. Una probabilidad para una variable continua es igual al área bajo la curva de densidad de dicha variable entre dos puntos.

Ejemplo 3.4.3 Glucosa en sangre

Consideremos el nivel de glucosa en sangre, en mg/dl, de un sujeto elegido aleatoriamente de la población descrita en el Ejemplo 3.4.2. Vimos en dicho ejemplo que el 42 % de los niveles de glucosa de la población están entre 100 mg/dl y 150 mg/dl. Por tanto, $\Pr\{100 \leq \text{nivel de glucosa} \leq 150\} = 0,42$.

Estamos modelando el nivel de glucosa en sangre como una variable continua, lo que significa que $\Pr\{\text{nivel de glucosa} = 100\} = 0$, como indicamos anteriormente, que $\Pr\{\text{nivel de glucosa} = 100\} = 0$ y $\Pr\{\text{nivel de glucosa} = 150\} = 0$. Por tanto,

$$\Pr\{100 \leq \text{nivel de glucosa} \leq 150\} = \Pr\{100 < \text{nivel de glucosa} < 150\} = 0,42$$



Ejemplo 3.4.4 Diámetros de árboles

El diámetro del tronco de un árbol es una variable importante en ingeniería forestal. La curva de densidad que se muestra en la Figura 3.4.5 representa la distribución de los diámetros (medidos 4,5 pies por encima de suelo) de una población de abetos Douglas de una edad de 30 años. En la figura se muestran varias áreas bajo la curva⁸. Considere el diámetro, en pulgadas, de un árbol escogido aleatoriamente. Entonces, por ejemplo, $\Pr\{4 < \text{diámetro} < 6\} = 0,33$. Si deseamos obtener la probabilidad de que un árbol escogido aleatoriamente tenga un diámetro mayor que 8 pulgadas, debemos sumar las últimas dos áreas bajo la curva de la Figura 3.4.3: $\Pr\{\text{diámetro} > 8\} = 0,12 + 0,07 = 0,19$. ■

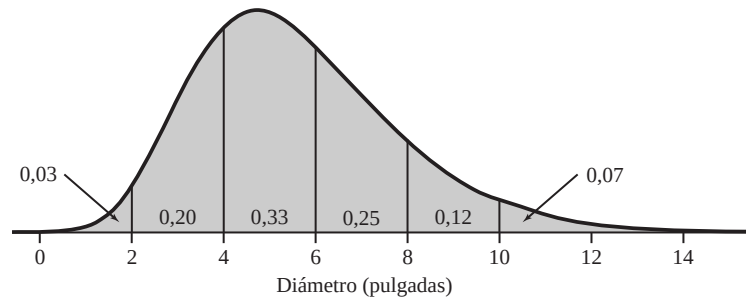


Figura 3.4.5 Diámetros de abetos Douglas de 30 años de edad

Ejercicios 3.4.1-3.4.4

3.4.1 Considere la curva de densidad que se muestra en la Figura 3.4.5, que representa la distribución de diámetros (medidos 4,5 pies por encima del suelo) de una población de abetos Douglas de 30 años de edad. La figura muestra varias áreas bajo la curva. ¿Qué porcentaje de los árboles tienen diámetros

- (a) entre cuatro pulgadas y 10 pulgadas?
- (b) menores que 4 pulgadas?
- (c) de más de 6 pulgadas?

3.4.2 Considere el diámetro de un abeto Douglas escogido aleatoriamente de la población representada por la curva de densidad que se muestra en la Figura 3.4.5. Calcule

- (a) $\Pr\{\text{diámetro} < 10\}$.
- (b) $\Pr\{\text{diámetro} > 4\}$.
- (c) $\Pr\{2 < \text{diámetro} < 8\}$.

3.4.3 En una cierta población del parásito *Trypanosoma*, las longitudes de los individuos se distribuyen como se indi-

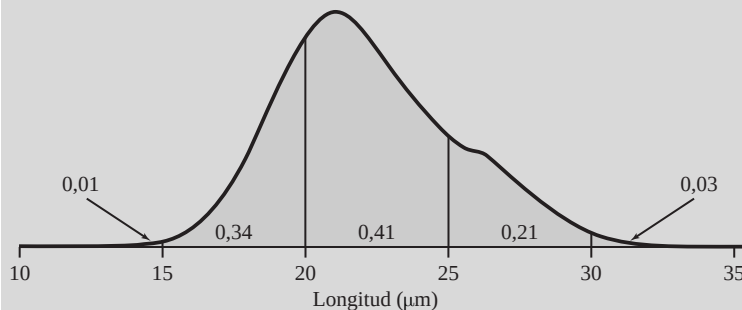
ca en la curva de densidad que se muestra a continuación. La figura muestra varias áreas bajo la curva⁹.

Considere la longitud de un tripanosoma escogida aleatoriamente de la población. Calcule

- (a) $\Pr\{20 < \text{longitud} < 30\}$.
- (b) $\Pr\{\text{longitud} > 20\}$.
- (c) $\Pr\{\text{longitud} < 20\}$.

3.4.4 Considere la distribución de longitudes de *Trypanosoma* definida por la curva de densidad del Ejercicio 3.4.3. Suponga que se toma una muestra de dos tripanosomas. ¿Cuál es la probabilidad de que

- (a) ambos tripanosomas tengan una longitud menor que 20 μm ?
- (b) el primer tripanosoma tenga una longitud menor que 20 μm y el segundo tripanosoma tenga una longitud mayor que 25 μm ?
- (c) exactamente uno de los tripanosomas tenga una longitud menor que 20 μm y un tripanosoma tenga una longitud mayor que 25 μm ?



3.5 Variables aleatorias

Una **variable aleatoria** es simplemente una variable que toma valores numéricos que dependen del resultado de una operación aleatoria. Los siguientes ejemplos ilustran esta idea.

Ejemplo 3.5.1 Dado

Consideremos la operación aleatoria de lanzar un dado. Sea la variable aleatoria Y que representa el número de puntos que se muestran. Los valores posibles de Y son $Y = 1, 2, 3, 4, 5$ o 6 . No conocemos el valor de Y hasta que hemos lanzado el dado. Si sabemos como está cargado el dado, entonces podemos calcular la probabilidad de que Y tenga un valor particular, por ejemplo $\Pr\{Y = 4\}$, o un conjunto particular de valores, como $\Pr\{2 \leq Y \leq 4\}$. Por ejemplo, si el dado está perfectamente equilibrado de forma que cualquiera de las caras es igualmente probable, entonces

$$\Pr\{Y = 4\} = \frac{1}{6} \approx 0,17$$

y

$$\Pr\{2 \leq Y \leq 4\} = \frac{3}{6} = 0,5$$

Ejemplo 3.5.2 Tamaño de familia

Supongamos que se escoge aleatoriamente una familia de una cierta población, y sea la variable aleatoria Y que indica el número de hijos de la familia elegida. Los valores posibles de Y son $0, 1, 2, 3, \dots$. La probabilidad de que Y tenga un valor concreto es igual al porcentaje de familias con ese número de hijos. Por ejemplo, si el 23 % de las familias tienen dos hijos, entonces

$$\Pr\{Y = 2\} = 0,23$$

Ejemplo 3.5.3 Medicaciones

Después de sufrir cirugía cardíaca, los pacientes generalmente reciben diversas medicaciones. Sea la variable aleatoria Y que indica el número de medicaciones que recibe un paciente después de recibir cirugía cardíaca. Si conocemos la distribución del número de medicaciones por paciente para toda la población, entonces podemos especificar la probabilidad de que Y tenga un cierto valor o esté entre un cierto intervalo de valores. Por ejemplo, si el 52 % de todos los pacientes reciben 2, 3, 4 o 5 medicaciones, entonces

$$\Pr\{2 \leq Y \leq 5\} = 0,52$$

Ejemplo 3.5.4 Alturas de hombres

Sea la variable aleatoria Y que indica la altura de un hombre elegido aleatoriamente de una cierta población. Si sabemos la distribución de las alturas de la población, entonces podemos especificar la probabilidad de que Y esté en un cierto intervalo. Por ejemplo, si el 46 % de los hombres tienen alturas entre 65,2 y 70,4 pulgadas, entonces

$$\Pr\{65,2 \leq Y \leq 70,4\} = 0,46$$

Todas las variables de los Ejemplos 3.5.1-3.5.3 son *variables aleatorias discretas*, porque en cada caso podemos hacer una lista con los posibles valores que puede tomar la variable. Por el contrario, la variable del Ejemplo 3.5.4, la altura, es una *variable aleatoria continua*: la altura, al menos en teoría, puede tomar cualquiera de un número infinito de valores en un intervalo. Por supuesto, cuando medimos y anotamos la altura a una persona, generalmente lo hacemos con una precisión de pulgada o de media pulgada. No obstante, podemos ver la altura verdadera como una variable continua. Para modelar las distribuciones de variables aleatorias continuas, como el nivel de glucosa en sangre o el diámetro de árboles, utilizaremos curvas de densidad, tal como se indicó en la Sección 3.4.

Media y varianza de una variable aleatoria

En el Capítulo 2 consideramos brevemente los conceptos de media poblacional y desviación típica poblacional. En el caso de una variable aleatoria discreta, podemos calcular la media poblacional y la desviación típica poblacional si conocemos la distribución de probabilidad de dicha variable aleatoria. Comenzaremos con la media.

La media de una variable aleatoria discreta se define como

$$\mu_Y = \sum y_i \Pr(Y = y_i)$$

siendo y_i los valores que toma la variable aleatoria, y la suma se realiza sobre todos los posibles valores.

La media de una variable aleatoria se conoce también como *valor esperado* y se expresa frecuentemente como $E(Y)$; es decir, $E(Y) = \mu_Y$.

Ejemplo 3.5.5 Vértebras de peces

En una cierta población de escorpiones de agua dulce, *Cottus rotheus*, la distribución del número de vértebras de la cola, Y , se muestra en la Tabla 3.5.1⁹.

Tabla 3.5.1 Distribución de vértebras	
Número de vértebras	Porcentaje de peces
20	3
21	51
22	40
23	6
Total	100

La media de Y es

$$\begin{aligned}
 \mu_Y &= 20 \times \Pr\{Y = 20\} + 21 \times \Pr\{Y = 21\} + 22 \times \Pr\{Y = 22\} + 23 \times \Pr\{Y = 23\} \\
 &= 20 \times 0,03 + 21 \times 0,51 + 22 \times 0,40 + 23 \times 0,06 \\
 &= 0,6 + 10,71 + 8,8 + 1,38 \\
 &= 21,49
 \end{aligned}$$

Ejemplo 3.5.6 Dado

Consideremos el lanzamiento de un dado perfectamente equilibrado, de forma que cada una de sus seis caras tiene la misma probabilidad de salir, y sea Y la variable aleatoria que representa el número de puntos que se muestran. El valor esperado, o media de Y es

$$E(Y) = \mu_Y = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{21}{6} = 3,5$$

Para calcular la desviación típica de una variable, se calcula primero la varianza, σ^2 , de dicha variable aleatoria y después se extrae la raíz cuadrada de la varianza, obteniéndose así la desviación típica, σ .

La varianza de una variable aleatoria discreta se define como

$$\sigma_Y^2 = \sum (y_i - \mu_Y)^2 \Pr(Y = y_i)$$

siendo y_i los valores que toma la variable aleatoria, y la suma se realiza sobre todos los posibles valores.

Escribiremos a menudo $\text{VAR}(Y)$ para indicar la varianza de Y .

Ejemplo 3.5.7 Vértebras de peces

Consideremos la distribución de vértebras de la Tabla 3.5.1. En el Ejemplo 3.5.5 obtuvimos que la media de Y es $\mu_Y = 21,49$. La varianza de Y es

$$\begin{aligned}\text{VAR}(Y) &= \sigma_Y^2 = (20 - 21,49)^2 \times \Pr\{Y = 20\} + (21 - 21,49)^2 \times \Pr\{Y = 21\} + (22 - 21,49)^2 \times \Pr\{Y = 22\} \\ &\quad + (23 - 21,49)^2 \times \Pr\{Y = 23\} \\ &= (-1,49)^2 \times 0,03 + (-0,49)^2 \times 0,51 + (0,51)^2 \times 0,40 + (1,51)^2 \times 0,06 \\ &= 2,2201 \times 0,03 + 0,2401 \times 0,51 + 0,2601 \times 0,40 + 2,2801 \times 0,06 \\ &= 0,066603 + 0,122451 + 0,10404 + 0,136806 \\ &= 0,4299\end{aligned}$$

La desviación típica de Y es $\sigma_Y = \sqrt{0,4299} \approx 0,6557$. ■

Ejemplo 3.5.8 Dado

En el Ejemplo 3.5.6 obtuvimos que el número medio obtenido al lanzar un dado equilibrado es 3,5 (es decir, $\mu_Y = 3,5$). La varianza del número obtenido al lanzar dicho dado es:

$$\begin{aligned}\sigma_Y^2 &= (1 - 3,5)^2 \times \Pr\{Y = 1\} + (2 - 3,5)^2 \times \Pr\{Y = 2\} + (3 - 3,5)^2 \times \Pr\{Y = 3\} + (4 - 3,5)^2 \times \Pr\{Y = 4\} \\ &\quad + (5 - 3,5)^2 \times \Pr\{Y = 5\} + (6 - 3,5)^2 \times \Pr\{Y = 6\} \\ &= (-2,5)^2 \times \frac{1}{6} + (-1,5)^2 \times \frac{1}{6} + (-0,5)^2 \times \frac{1}{6} + (0,5)^2 \times \frac{1}{6} + (1,5)^2 \times \frac{1}{6} + (2,5)^2 \times \frac{1}{6} + (1,5)^2 \times \frac{1}{6} + (2,5)^2 \times \frac{1}{6} \\ &= (6,25) \times \frac{1}{6} + (2,25) \times \frac{1}{6} + (0,25) \times \frac{1}{6} + (0,25) \times \frac{1}{6} + (2,25) \times \frac{1}{6} + (6,25) \times \frac{1}{6} \\ &= 17,5 \times \frac{1}{6} \\ &\approx 2,9167\end{aligned}$$

La desviación típica de Y es $\sigma_Y = \sqrt{2,9167} \approx 1,708$. ■

Las definiciones anteriores son adecuadas para variables aleatorias discretas. Existen definiciones análogas para variables aleatorias continuas, pero en ellas aparecen integrales y no las presentaremos aquí.

Suma y diferencia de variables aleatorias (opcional)

Si sumamos dos variables aleatorias, tiene sentido que sus medias también se sumen. Asimismo, si creamos una nueva variable aleatoria restando dos variables aleatorias entonces para obtener la media de la nueva variable se restan las medias de cada una de ellas. Si multiplicamos una variable aleatoria por una constante (por ejemplo, si estamos convirtiendo pies en pulgadas, de forma que multiplicamos por 12), entonces se multiplica la media de la variable aleatoria por la misma constante. Si se suma una constante a una variable aleatoria, entonces se suma dicha constante a la media.

Las siguientes reglas resumen la situación:

Reglas para las medias de variables aleatorias

Regla (1). Si X e Y son dos variables aleatorias independientes entre sí, entonces $\mu_{X+Y} = \mu_X + \mu_Y$.

$$\mu_{X-Y} = \mu_X - \mu_Y$$

Regla (2). Si Y es una variable aleatoria y a y b son constantes, entonces

$$\mu_{a+bY} = a + b\mu_Y.$$

Ejemplo 3.5.9 Temperatura

La temperatura media en verano, μ_Y , en una ciudad es de 81 °F. Para convertir °F en °C, utilizamos la fórmula $^{\circ}\text{C} = (^{\circ}\text{F} - 32) \times (5/9)$ o $^{\circ}\text{C} = (5/9) \times ^{\circ}\text{F} - (5/9) \times 32$. Por tanto, la media en grados Celsius es $(5/9) \times (81) - (5/9) \times 32 = 45 - 17,78 = 27,22$. ■

Considerar las desviaciones típicas de funciones de variables aleatorias es un poco más complicado. Trabajaremos primero con la varianza y después tomaremos la raíz cuadrada, al final, para obtener la desviación típica deseada. Si se *multiplica* una variable aleatoria por una constante (por ejemplo, si estamos convirtiendo pulgadas a centímetros multiplicando por 2,54), entonces la varianza se multiplica por el cuadrado de la constante. Esto tiene el efecto de multiplicar la desviación típica por el valor absoluto de la constante. Si se *suma* una constante a una variable aleatoria, entonces no se está cambiando la dispersión relativa de la distribución, por lo que la varianza no cambia.

Ejemplo 3.5.10 Pies a pulgadas

Sea Y la altura, en pies, de una persona de una determinada población. Supongamos que la desviación típica de Y es $\sigma_Y = 0,35$ (pies). Si deseamos convertir pies en pulgadas, podemos definir una nueva variable X como $X = 12Y$. La varianza de Y es $0,35^2$ (el cuadrado de la desviación típica). La varianza de X es $12^2 \times 0,35^2$, lo que significa que la desviación típica de X es $\sigma_X = 12 \times 0,35 = 4,2$ (pulgadas). ■

Si sumamos dos variables aleatorias *que sean independientes entre sí*, entonces sus varianzas se suman*. Además, si restamos dos variables aleatorias *que sean independientes entre sí*, sus varianzas se suman. Si deseamos calcular la desviación típica de la suma (o diferencia) de dos variables aleatorias independientes, calcularemos primero la varianza de la suma (o diferencia), y tomaremos después la raíz cuadrada, obteniendo así la desviación típica de la suma (o diferencia).

Ejemplo 3.5.11 Masa

Consideremos la obtención de la masa de un cilindro graduado de 10 ml. Si se realizan varias medidas, utilizando una balanza analítica, entonces en teoría esperaríamos que las medidas fueran idénticas. En la realidad, sin embargo, las lecturas variarán de una medida a la siguiente. Supongamos que una determinada balanza produce lecturas que tienen una desviación típica de 0,03 g. Sea X el valor de la lectura realizada utilizando esta balanza. Supongamos que una segunda balanza produce lecturas que tienen una desviación típica de 0,04 g. Sea Y el valor de una lectura realizada utilizando esta segunda balanza¹⁰.

Si utilizamos las dos balanzas para medir la masa de un cilindro graduado, podríamos estar interesados, por ejemplo, en la diferencia, $X - Y$, de las dos medidas. La desviación típica de $X - Y$ es positiva. Para obtener la desviación típica de $X - Y$, obtenemos primero la varianza de la diferencia. La varianza de X es $0,03^2$ y la varianza de Y es $0,04^2$. La varianza de la diferencia es $0,03^2 + 0,04^2 = 0,0025$. La desviación típica de $X - Y$ es la raíz cuadrada de 0,0025, que es 0,005. ■

Las siguientes reglas resumen la situación para la varianza:

Reglas para las varianzas de variables aleatorias

Regla (3). Si Y es una variable aleatoria y a y b constantes, entonces $\sigma_{a+bY}^2 = b^2\sigma_Y^2$.

Regla (4). Si X y Y son dos variables aleatorias *independientes*, entonces

$$\begin{aligned}\sigma_{X+Y}^2 &= \sigma_X^2 + \sigma_Y^2 \\ \sigma_{X-Y}^2 &= \sigma_X^2 + \sigma_Y^2\end{aligned}$$

* Si sumamos dos variables aleatorias que no sean independientes entre sí, entonces la varianza de la suma depende del grado de dependencia entre las variables. Para tomar un caso extremo, supongamos que una de las variables es la otra cambiada de signo. Entonces la suma de las dos variables aleatorias será siempre cero, de forma que la varianza de la suma será cero. Esto es bastante diferente de lo que obtendríamos sumando las dos varianzas. Como otro ejemplo, supongamos que Y es el número de preguntas correctas en un examen de 20 preguntas, y X el número de preguntas incorrectas. Entonces $Y + X$ es siempre igual a 20, de forma que la suma no presenta variabilidad en absoluto. Entonces, la varianza de $Y + X$ es siempre cero, incluso aunque la varianza de Y y la varianza de X sean ambas positivas.

Ejercicios 3.5.1-3.5.8

3.5.1 En una cierta población de estorninos europeos hay 5.000 nidos con crías. La distribución del tamaño de la nidada (número de crías en un nido) se muestra en la tabla siguiente¹¹.

Tamaño de la nidada	Frecuencia (n.º de nidadas)
1	90
2	230
3	610
4	1.400
5	1.760
6	750
7	130
8	26
9	3
10	1
Total	5.000

Suponga que se escoge aleatoriamente una de las nidadas, y sea Y el tamaño de la misma. Calcule

- (a) $\Pr\{Y = 3\}$.
- (b) $\Pr\{Y \geq 7\}$.
- (c) $\Pr\{4 \leq Y \leq 6\}$.

3.5.2 En la población de estorninos del Ejercicio 3.5.1 hay un total de 22.435 crías en todas las nidadas. (Hay 90 crías de nidadas de tamaño 1, 460 de nidadas de tamaño 2, etc.). Supongamos que se escoge aleatoriamente una de las crías, y sea Y' el tamaño de la nidada escogida.

- (a) Calcule $\Pr\{Y' = 3\}$.
- (b) Calcule $\Pr\{Y' \geq 7\}$.
- (c) Explique por qué escoger aleatoriamente una cría y después observar su nidada no es equivalente a escoger aleatoriamente una nidada. Su explicación debería demostrar por qué la respuesta al apartado (b) es mayor que la respuesta al apartado (b) del Ejercicio 3.5.1.

3.5.3 Calcule la media, μ_Y , de la variable aleatoria Y del Ejercicio 3.5.1.

3.5.4 Considere una población de moscas de la fruta, *Drosophila melanogaster*, en la que el 30 % de los individuos son negros debido a una mutación, mientras que el 70 % de los individuos mantiene el color normal gris del cuerpo. Supongamos que se escogen aleatoriamente tres moscas de la población. Sea Y el número de moscas negras en las tres. Entonces, la distribución de probabilidad de Y está dada por la siguiente tabla:

Y (n.º de negras)	Probabilidad
0	0,343
1	0,441
2	0,189
3	0,027
Total	1,000

- (a) Calcule $\Pr\{Y \geq 2\}$.
- (b) Calcule $\Pr\{Y \leq 2\}$.

3.5.5 Calcule la media, μ_Y , de la variable aleatoria Y del Ejercicio 3.5.4.

3.5.6 Calcule la desviación típica, σ_Y , de la variable aleatoria Y del Ejercicio 3.5.4.

3.5.7 Se preguntó a un grupo de estudiantes universitarios cuántas veces habían visitado al dentista en el último año¹². La distribución de probabilidad de Y , el número de visitas, está dada por la siguiente tabla:

Y (n.º de visitas)	Probabilidad
0	0,15
1	0,50
2	0,35
Total	1,00

Calcule la media, μ_Y , del número de visitas.

3.5.8 Calcule la desviación típica, σ_Y , de la variable aleatoria Y del Ejercicio 3.5.7.

3.6 La distribución binomial

Para profundizar en las nociones de probabilidad y variables aleatorias, consideraremos ahora un tipo especial de variable aleatoria, la **binomial**. La distribución de una variable aleatoria binomial es una distribución de probabilidad asociada a una clase especial de operación aleatoria. Dicha operación aleatoria se define en términos de una serie de condiciones que se denominan modelo de pruebas independientes.

El modelo de pruebas independientes

El **modelo de pruebas independientes** se relaciona con una secuencia de «pruebas» aleatorias. Cada prueba solo puede tener dos posibles resultados, que se denominan «éxito» y «fracaso». La probabilidad de éxito de cada prueba individual se indica mediante la letra p y se supone constante de una prueba a la siguiente. Además, se requiere que las pruebas sean independientes, lo que significa que la probabilidad de éxito o fracaso en cada prueba no depende del resultado de ninguna otra prueba. El número total de pruebas se denomina n . Estas condiciones se resumen en la siguiente definición del modelo.

Modelo de pruebas independientes

Se realiza una serie de n pruebas independientes. Cada prueba puede resultar un éxito o un fracaso. La probabilidad de éxito es igual a la misma cantidad, p , en cada prueba, independientemente del resultado de las otras pruebas.

Los ejemplos siguientes ilustran situaciones que se pueden describir mediante el modelo de pruebas independientes.

Ejemplo 3.6.1 Albinismo

Si se casan dos personas portadoras del gen del albinismo, cada uno de sus hijos tiene una probabilidad de $1/4$ de ser albino. La probabilidad de que el segundo hijo sea albino es la misma ($1/4$), independientemente de si el primer hijo fue o no fue albino. De forma similar, el resultado del tercer hijo es independiente de los otros dos, y así sucesivamente. Utilizando las etiquetas «éxito» para el caso de albino y «fracaso» para el caso de no albino, el modelo de pruebas independientes se puede aplicar con $p = 1/4$ y $n =$ el número de hijos de la familia. ■

Ejemplo 3.6.2 Gatos mutantes

En un estudio sobre gatos realizado en Omaha, se encontró que el 37 % tenían un cierto rasgo mutante¹³. Supongamos que el 37 % de todos los gatos tienen esta característica mutante y que la muestra aleatoria de gatos se extrae de la población. A medida que se extrae cada gato de la muestra, la probabilidad de que sea mutante es de 0,37. Esta probabilidad es la misma para cada gato elegido, independientemente de los otros gatos, porque el porcentaje de mutantes en la población grande sigue siendo igual a 0,37 aunque se extraigan unos pocos gatos de la población. Etiquetando como «éxito» al mutante y como «fracaso» al no mutante, se puede aplicar el modelo de pruebas independientes con $p = 0,37$ y $n =$ el tamaño de la muestra. ■

Un ejemplo de distribución binomial

La distribución binomial especifica las probabilidades de diversos números de éxitos y fracasos cuando la operación aleatoria básica consiste en n pruebas independientes. Antes de presentar la fórmula general de la distribución binomial, consideremos un ejemplo simple.

Ejemplo 3.6.3 Albinismo

Supongamos que se casan dos personas portadoras del gen del albinismo (véase el Ejemplo 3.6.1) y que tienen dos hijos. Entonces, la probabilidad de que ambos sean albinos es

$$\Pr\{\text{ambos hijos son albinos}\} = \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16}$$

La razón de ese valor de la probabilidad se puede ver considerando la interpretación de la probabilidad como frecuencia relativa. De un conjunto muy grande de familias de este tipo, $\frac{1}{4}$ tendrían el primer hijo albino. Además, $\frac{1}{4}$ de estas tendrían el segundo hijo albino. Por tanto, $\frac{1}{4}$ de $\frac{1}{4}$, o $\frac{1}{16}$ de todas las parejas tendrían ambos hijos albinos. Un tipo de razonamiento similar permite demostrar que la probabilidad de que ambos hijos no sean albinos es

$$\Pr\{\text{ambos hijos no son albinos}\} = \left(\frac{3}{4}\right)\left(\frac{3}{4}\right) = \frac{9}{16}$$

Una vuelta de tuerca más: consideremos la probabilidad de que un hijo sea albino y el otro no. Hay dos posibles formas en las que esto puede suceder:

$$\Pr\{\text{el primer hijo es albino, el segundo no}\} = \left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{3}{16}$$

$$\Pr\{\text{el primer hijo no es albino, el segundo sí}\} = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right) = \frac{3}{16}$$

Para ver cómo combinar estas probabilidades, consideremos de nuevo la interpretación de la probabilidad como frecuencia relativa. De un conjunto muy grande de familias de este tipo con dos hijos, la fracción de familias con un hijo albino y el otro no sería el total de las dos posibilidades, o

$$\left(\frac{3}{16}\right) + \left(\frac{3}{16}\right) = \frac{6}{16}$$

Por tanto, la probabilidad correspondiente es

$$\Pr\{\text{un primer hijo es albino, el otro no}\} = \frac{6}{16}$$

Otra forma de ver esto es utilizar un árbol de probabilidad. La primera bifurcación del árbol representa el nacimiento del primer hijo. La segunda bifurcación representa el nacimiento del segundo hijo. Los cuatro posibles resultados y sus probabilidades asociadas se muestran en la Figura 3.6.1. Dichas probabilidades se recogen en la Tabla 3.6.1. ■

La distribución de probabilidades de la Tabla 3.6.1 se denomina distribución binomial con $p = \frac{1}{4}$ y $n = 2$. Nótese que las probabilidades suman 1. Esto tiene sentido porque se han tenido en cuenta todas las posibilidades: Esperamos que $\frac{9}{16}$ de las familias no tengan hijos albinos, que $\frac{6}{16}$ tengan un hijo albino y que $\frac{1}{16}$ tengan dos hijos albinos. No existen otras posibilidades para una familia de dos hijos. El número de hijos albinos de los dos hijos considerados es un ejemplo de variable aleatoria binomial. Una **variable aleatoria binomial** es una variable que satisface las siguientes cuatro condiciones, abreviadas como **BInM**:

Tabla 3.6.1 Distribución de probabilidad del número de hijos albinos		
Número de		Probabilidad
Albino	No albino	
0	2	$\frac{9}{16}$
1	1	$\frac{6}{16}$
2	0	$\frac{1}{16}$
		1

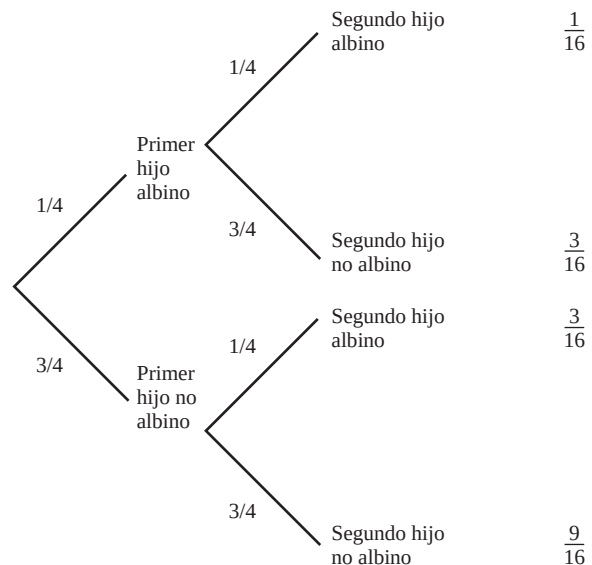


Figura 3.6.1 Árbol de probabilidad para el albinismo entre dos hijos de portadores del gen del albinismo.

Resultados Binarios: solo hay dos posibles resultados para cada prueba (éxito o fracaso).

Pruebas Independientes: los resultados de las pruebas son independientes entre sí.

n es fijo: el número de pruebas n , se fija al principio.

Mismo valor de p : la probabilidad de éxito en una sola prueba es la misma en todas las pruebas.

La fórmula de la distribución binomial

Existe una fórmula general que se puede utilizar para calcular probabilidades asociadas a una variable aleatoria binomial para cualquier valor de n y de p . Esta fórmula se puede demostrar utilizando una lógica similar a la del Ejemplo 3.6.3. (La fórmula se considera después en el Apéndice 3.1). Dicha fórmula se presenta en el recuadro siguiente.

La fórmula de la distribución binomial

Dada una variable aleatoria binomial Y , la probabilidad de que en n pruebas ocurran j éxitos (y $n - j$ fallos) viene dada por la siguiente fórmula:

$$\Pr\{j \text{ éxitos}\} = \Pr\{Y = j\} = {}_nC_j p^j (1 - p)^{n-j}$$

La cantidad ${}_nC_j$ que aparece en la fórmula se denomina **coeficiente binomial**. Cada coeficiente binomial es un entero que depende de n y j . Los valores de los coeficientes binomiales se dan en la Tabla 2 al final del libro y se pueden calcular mediante la fórmula

$${}_nC_j = \frac{n!}{j!(n-j)!}$$

Donde $x!$ («factorial de x ») está definido para cualquier entero positivo x como

$$x! = x(x-1)(x-2)\dots(2)(1)$$

y $0! = 1$. Para más detalles, véase el Apéndice 3.1.

Por ejemplo, para $n = 5$, los coeficientes binomiales son los siguientes:

$$\begin{array}{rcccccc} j: & 0 & 1 & 2 & 3 & 4 & 5 \\ {}_5C_j: & 1 & 5 & 10 & 10 & 5 & 1 \end{array}$$

Por tanto, para $n = 5$, las probabilidades binomiales son las indicadas en la Tabla 3.6.2. Nótese el patrón que aparece en dicha tabla: Las potencias de p ascienden (0, 1, 2, 3, 4, 5) y las potencias de $(1 - p)$ descienden (5, 4, 3, 2, 1, 0). (Al utilizar la fórmula de la distribución binomial, recuérdese que $x^0 = 1$ para todo x distinto de cero).

Tabla 3.6.2 Probabilidades binomiales para $n = 5$		
Número de		
Éxitos j	Fracasos $n - j$	Probabilidad
0	5	$1p^0(1 - p)^5$
1	4	$5p^1(1 - p)^4$
2	3	$10p^2(1 - p)^3$
3	2	$10p^3(1 - p)^2$
4	1	$5p^4(1 - p)^1$
5	0	$1p^5(1 - p)^0$

El ejemplo siguiente muestra una aplicación específica de la distribución binomial para $n = 5$.

Ejemplo 3.6.4 Gatos mutantes

Supongamos que extraemos una muestra aleatoria de cinco individuos de una población grande en la que el 37 % de los individuos son mutantes (como en el Ejemplo 3.6.2). Las probabilidades de las diversas muestras posibles están dadas entonces por la fórmula de la distribución binomial con $n = 5$ y $p = 0,37$. Los resultados se muestran en la Tabla 3.6.3. Por ejemplo, la probabilidad de que una muestra contenga 2 mutantes y 3 no mutantes, es

$$10(0,37)^2(0,63)^3 \approx 0,34$$

Tabla 3.6.3 Distribución binomial con $n = 5$ y $p = 0,37$		
Número de		Probabilidad
Mutantes	No mutantes	
0	5	0,10
1	4	0,29
2	3	0,34
3	2	0,20
4	1	0,06
5	0	0,01
		<u>1,00</u>

Por tanto, $\Pr\{Y = 3\} \approx 0,34$. Esto significa que aproximadamente el 34 % de las muestras aleatorias de tamaño 5 contendrán dos mutantes y tres no mutantes.

Nótese que las probabilidades de la Tabla 3.6.3 suman 1. Las probabilidades de una distribución de probabilidad deben siempre sumar 1, porque tienen en cuenta el 100 % de las posibilidades. ■

La distribución binomial de la Tabla 3.6.3 se muestra gráficamente en la Figura 3.6.2. Los picos de la gráfica resaltan el hecho de que la distribución de probabilidad es discreta.



Figura 3.6.2 Distribución binomial con $n = 5$ y $p = 0,37$

Observación Al aplicar el modelo de pruebas independientes y la distribución binomial, las etiquetas «éxito» y «fracaso» se asignan arbitrariamente. Por ejemplo, en el Ejemplo 3.6.4, podríamos asignar «éxito» = «mutante» y $p = 0,37$. Pero, de forma alternativa, podríamos asignar «éxito» = «no mutante», con probabilidad $p = 0,63$. Cualquier asignación es correcta; solo es necesario ser coherente.

Notas sobre la Tabla 2 Es conveniente notar algunos aspectos de la Tabla 2 (ver al final del libro):

- El primer y el último valor de cada fila son iguales a 1. Esto es cierto para cualquier fila. Es decir, ${}_nC_0 = 1$ y ${}_nC_n = 1$ para cualquier valor de n .
- Todas las filas de la tabla son simétricas, es decir, ${}_nC_j$ y ${}_nC_{n-j}$ son iguales.

- (c) Las últimas filas de la tabla se dejan incompletas para ahorrar espacio, pero se pueden completar fácilmente utilizando la simetría de los coeficientes ${}_nC_j$. Si se necesita conocer ${}_nC_j$ se puede buscar el valor ${}_nC_{n-j}$ en la Tabla 2. Por ejemplo, sea $n = 18$. Si se desea conocer ${}_{18}C_{15}$ hay que buscar ${}_{18}C_3$. Tanto ${}_{18}C_3$ como ${}_{18}C_{15}$ valen 816.

Nota sobre el cómputo La tecnología de computadores y calculadoras hacen bastante fácil manejar la fórmula de la distribución binomial para valores pequeños o moderados de n . Para valores grandes de n , el uso de la fórmula binomial se hace tedioso e incluso un computador puede tener problemas cuando se le pide calcular una fórmula binomial en ese caso. Sin embargo, la fórmula binomial se puede aproximar por otros métodos. Uno de ellos será presentado en la Sección opcional 5.5.

Algunas veces una pregunta sobre una distribución binomial implica combinar dos o más resultados posibles. El ejemplo siguiente ilustra esta idea.

Ejemplo 3.6.5 Muestreo de moscas de la fruta

En una población grande de *Drosophila*, el 30 % de las moscas son negras (B) y el 70 % son grises (G). Supongamos que se escogen aleatoriamente dos moscas de la población (como en el Ejemplo 3.2.3). La distribución binomial con $n = 2$ y $p = 0,3$ proporciona las probabilidades de los posibles resultados como se muestra en la Tabla 3.6.4. (El uso de la fórmula binomial concuerda con los resultados dados por el árbol de probabilidad que se muestra en la Figura 3.2.3).

Tabla 3.6.4		
Composición de la muestra	Y	Probabilidad
Ambas G	0	0,49
Una B, una G	1	0,42
Ambas B	2	0,09
		1,00

Sea E el suceso de que ambas moscas son del mismo color. Entonces E puede suceder de dos formas: ambas moscas son grises o ambas son negras. Para calcular la probabilidad de E , consideremos lo que podría ocurrir si repitiéramos el procedimiento de muestreo muchas veces: el 49 % de las muestras tendrían ambas moscas grises, y el 9 % tendrían ambas moscas negras. Por consiguiente, el porcentaje de muestras con ambas moscas del mismo color sería $49 \% + 9 \% = 58 \%$. Por tanto, hemos demostrado que la probabilidad de E es

$$\Pr\{E\} = 0,58$$

tal como indicamos en el Ejemplo 3.2.3. ■

Siempre que un suceso E pueda suceder de dos o más formas mutuamente exclusivas, un razonamiento como el del Ejemplo 3.6.5 se puede usar para obtener $\Pr\{E\}$.

Ejemplo 3.6.6 Grupo sanguíneo

En Estados Unidos, el 85 % de la población tiene sangre Rh positivo. Supongamos que tomamos una muestra aleatoria de seis personas y contamos el número que tiene el Rh positivo. El modelo binomial se puede aplicar aquí, ya que las condiciones BInM se cumplen: cada prueba produce un resultado binario (Rh positivo o Rh negativo), las pruebas son independientes (debido al muestreo aleatorio), n tiene un valor fijo de 6, y todas las personas tienen la misma probabilidad de tener sangre Rh positivo ($p = 0,85$).

Sea Y el número de personas de las 6 que tienen sangre Rh positivo. Las probabilidades de los posibles valores de Y están dadas por la fórmula de la distribución binomial con $n = 6$ y $p = 0,85$. Los resultados se muestran en la Tabla 3.6.5. Por ejemplo, la probabilidad de que $Y = 4$ es

$${}_6C_4(0,85)^4(0,15)^2 \approx 15(0,522)(0,0225) \approx 0,1762$$

Si deseamos calcular la probabilidad de que al menos 4 personas (de las 6 muestreadas) tengan sangre Rh positivo, necesitamos calcular $\Pr\{Y \geq 4\} = \Pr\{Y = 4\} + \Pr\{Y = 5\} + \Pr\{Y = 6\} = 0,1762 + 0,3993 + 0,3771 = 0,9526$. Esto significa que la probabilidad de obtener al menos 4 personas con sangre Rh positivo en una muestra de tamaño 6 es 0,9526. ■

Tabla 3.6.5 Distribución binomial con $n = 6$ y $p = 0,85$	
Número de éxitos	Probabilidad
0	<0,0001
1	0,0004
2	0,0055
3	0,0415
4	0,1762
5	0,3993
6	0,3771
	1

En algunos problemas, es más fácil calcular la probabilidad de que un suceso *no ocurra* en vez de calcular la probabilidad de que el suceso ocurra. Para resolver estos problemas utilizamos el hecho de que la probabilidad de que un suceso ocurra es uno menos la probabilidad de que el suceso no ocurra: $\Pr\{E\} = 1 - \Pr\{E \text{ no ocurra}\}$. A continuación sigue un ejemplo.

Ejemplo 3.6.7 Grupo sanguíneo

Como en el Ejemplo 3.6.6, sea Y la variable que indica el número de personas, de las 6, con sangre Rh positivo. Supongamos que deseamos calcular la probabilidad de que Y sea menor que 6 (es decir, la probabilidad de que haya *al menos* 1 persona en la muestra que tenga sangre Rh *negativo*). Podríamos calcular esa probabilidad directamente como $\Pr\{Y = 0\} + \Pr\{Y = 1\} + \dots + \Pr\{Y = 5\}$. Sin embargo, es más fácil calcular $\Pr\{Y \neq 6\}$, y restar este valor de 1:

$$\Pr\{Y < 6\} = 1 - \Pr\{Y = 6\} = 1 - 0,3771 = 0,6229$$

Media y desviación típica de una variable aleatoria binomial

Si lanzamos 10 veces una moneda equilibrada, entonces esperamos obtener en promedio cinco caras. Esto es un ejemplo de la regla general: *para una variable aleatoria binomial, la media (es decir, el número medio de éxitos), es igual a np* . Esto es un hecho intuitivo: la probabilidad de éxito en cada prueba es p , por lo que si realizamos n pruebas, entonces np es el número esperado de éxitos. En el Apéndice 3.2 demostramos que este resultado es consistente con la regla dada en la Sección 3.5 para obtener la media de la suma de variables aleatorias. La desviación típica de una variable aleatoria binomial está dada por $\sqrt{np(1-p)}$. Esta fórmula no resulta intuitivamente clara, y en el Apéndice 3.2 se presenta la obtención de este resultado. Para el ejemplo de lanzar una moneda 10 veces, la desviación típica del número de caras es

$$\sqrt{10 \times 0,5 \times 0,5} = \sqrt{2,5} \approx 1,58$$

Ejemplo 3.6.8 Grupo sanguíneo

Como se presentó en el Ejemplo 3.6.6, si Y indica el número de personas con sangre Rh positivo en una muestra de tamaño 6, entonces se puede utilizar un modelo binomial para obtener las probabilidades asociadas con Y . El valor más probable de Y es 5 (que tiene una probabilidad de 0,3993). El valor medio de Y es $6 \times 0,85 = 5,1$, lo que

significa que si tomáramos muchas muestras, todas ellas de tamaño 6, y contáramos el número de personas con Rh positivo en cada muestra, y después promediáramos dichos números, esperaríamos obtener un resultado de 5,1. La desviación típica de esos números es $\sqrt{6 \times 0,85 \times 0,15} \approx 0,87$. ■

Aplicabilidad de la distribución binomial

Diversos procedimientos estadísticos se basan en la distribución binomial. Estudiaremos algunos de estos procedimientos en capítulos posteriores. Por supuesto, la distribución binomial es aplicable solo en experimentos donde se cumplen las condiciones BInM en la situación biológica real. Comentaremos brevemente algunos aspectos de estas condiciones.

Aplicación al muestreo La aplicación más importante del modelo de pruebas independientes y de la distribución binomial es la de describir el muestreo aleatorio de una población donde la variable observada es dicotómica (es decir, una variable categórica con dos categorías, como negro y gris en el Ejemplo 3.6.5). Esta aplicación es válida si el tamaño de la muestra es una fracción despreciable del tamaño de la población, de forma que la composición de la población no se altera apreciablemente por la extracción de los individuos de la muestra (de forma que el apartado M de las condiciones BInM se cumple: la probabilidad de éxito no cambia de una prueba a otra). Sin embargo, si la muestra *no* es una parte pequeña despreciable de la población, entonces la composición de dicha población se puede alterar por el proceso de muestreo, de forma que las «pruebas» utilizadas para formar la muestra no son independientes y la probabilidad de un éxito cambia a medida que el muestreo progresa. En este caso, las probabilidades dadas por una binomial no son correctas. En la mayoría de los estudios biológicos, la población es tan grande que esta dificultad no aparece.

Contagio En algunas aplicaciones el fenómeno del contagio puede invalidar la condición de independencia entre pruebas. Sigue a continuación un ejemplo.

Ejemplo 3.6.9 Varicela

Consideremos la aparición de varicela en niños. Cada niño de una familia se puede clasificar de acuerdo a si tuvo varicela durante un cierto año. Se puede decir que cada niño constituye una «prueba» y que «éxito» es haber tenido varicela durante el año, pero las pruebas *no* son independientes debido a que la probabilidad de que un niño en particular enferme de varicela depende de si sus hermanos enfermaron de varicela. Como un ejemplo concreto, consideremos una familia con cinco niños, y supongamos que la probabilidad de que cada niño enferme de varicela durante el año es igual a 0,10. La distribución binomial da la probabilidad de que los cinco niños enfermen de varicela, y su valor es

$$\Pr\{5 \text{ niños con varicela}\} = (0,10)^5 = 0,00001$$

Sin embargo, la respuesta no es correcta. Debido al contagio, la probabilidad correcta sería mucho mayor. Habría muchas familias en las que un niño enfermaría de varicela y después los otros cuatro niños también, por contagiarse del primero, de forma que los cinco niños enfermarían de varicela. ■

Ejercicios 3.6.1-3.6.10

3.6.1 Las semillas de los guisantes (*Pisum sativum*) son amarillas o verdes. Un cierto cruce entre plantas de guisantes produce una progenie en la razón de 3 amarillos: 1 verde¹⁴. Si se escogen aleatoriamente cuatro descendientes de ese cruce y se examinan, calcule la probabilidad de que:

- (a) Tres sean amarillos y uno sea gris.
- (b) Los cuatro sean amarillos.
- (c) Los cuatro sean del mismo color.

3.6.2 En Estados Unidos, el 42 % de la población tiene grupo sanguíneo A. Considere extraer una muestra de tamaño 4. Sea Y la variable que indica el número de personas de la muestra con grupo sanguíneo A. Calcule

- (a) $\Pr\{Y = 0\}$.
- (b) $\Pr\{Y = 1\}$.
- (c) $\Pr\{Y = 2\}$.
- (d) $\Pr\{0 \leq Y \leq 2\}$.
- (e) $\Pr\{0 < Y \leq 2\}$.

3.6.3 Un cierto tratamiento con un medicamento cura el 90 % de los casos de anquilostoma en niños¹⁵. Suponga que 20 niños que tienen anquilostoma van a ser tratados, y que los niños se pueden considerar como una muestra aleatoria de la población. Calcule la probabilidad de que:

- (a) Los 20 se curen.
- (b) Se curen todos excepto 1.
- (c) Se curen exactamente 18.
- (d) Se cure exactamente el 90 %.

3.6.4 La concha del caracol de tierra *Limacolaria marteniana* presenta dos posibles formas de color: a rayas o pálida. En una cierta población de estos caracoles, el 60 % de los individuos tienen conchas a rayas¹⁶. Suponga que se extrae una muestra aleatoria de 10 caracoles de esta población. Calcule la probabilidad de que el porcentaje de caracoles con concha a rayas en la muestra sea

- (a) 50 %.
- (b) 60 %.
- (c) 70 %.

3.6.5 Considere la extracción de una muestra de tamaño 10 de la población de caracoles del Ejercicio 3.6.4.

- (a) ¿Cuál es el número medio de caracoles con concha a rayas?
- (b) ¿Cuál es la desviación típica del número de caracoles con concha a rayas?

3.6.6 La proporción de sexos de niños humanos recién nacidos es aproximadamente 105 hombres: 100 mujeres¹⁷. Si se escogen cuatro niños aleatoriamente, calcule la probabilidad de que:

- (a) Dos sean hombres y dos sean mujeres.
- (b) Los cuatro sean hombres.
- (c) Los cuatro sean del mismo sexo.

3.6.7 Construya un experimento binomial (diferente de los ejemplos presentados en este libro) y un problema que tenga el siguiente resultado: ${}_7C_3(0,8)^3(0,2)^5$.

3.6.8 El neuroblastoma es una enfermedad rara y seria, pero tratable. Se ha desarrollado una prueba de orina, la prueba VMA, que produce un diagnóstico positivo en aproximadamente el 70 % de casos de neuroblastoma¹⁸. Se ha propuesto que esta prueba se utilice para la revisión a gran escala de niños. Supongamos que se van a revisar 300.000 niños, de los cuales ocho tienen la enfermedad. Se desea saber si la prueba detecta o no la enfermedad en los ocho niños que la tienen. Calcule la probabilidad de que:

- (a) Se detecten los ocho casos.
- (b) Solo se deje un caso sin detectar.
- (c) Se dejen sin detectar dos o más casos. [Sugerencia: utilice los apartados (a) y (b) para responder al apartado (c)].

3.6.9 Si se casan dos portadores del gen del albinismo, cada uno de sus hijos tiene una probabilidad de 1/4 de ser albino (véase el Ejemplo 3.6.1). Si una pareja de este tipo tiene seis niños, calcule la probabilidad de que

- (a) Ninguno de ellos sea albino.
- (b) Al menos uno de ellos sea albino. [Sugerencia: utilice el apartado (a) para responder al apartado (b). Advierta que «al menos uno» significa «uno o más»].

3.6.10 La intoxicación de niños por plomo es un problema de salud pública en Estados Unidos. En una cierta población, 1 niño de cada 8 tiene un alto nivel de plomo en sangre (definido como 30 $\mu\text{g}/\text{dl}$ o más)¹⁹. En un grupo de 16 niños escogidos aleatoriamente de esta población, calcule la probabilidad de que:

- (a) Ninguno tenga un nivel alto de plomo en sangre.
- (b) 1 tenga un nivel alto de plomo en sangre.
- (c) 2 tengan un nivel alto de plomo en sangre.
- (d) 3 o más tengan un nivel alto de plomo en sangre. [Sugerencia: utilice los apartados (a)-(c) para responder al apartado (d)].

3.7 Ajuste de una distribución binomial a los datos (opcional)

Ocasionalmente, es posible obtener datos que permitan una comprobación directa de la aplicabilidad de la distribución binomial. En el ejemplo siguiente se describe un caso así.

Ejemplo 3.7.1 Sexos de niños

En un estudio clásico de la proporción de sexos en humanos, se clasificaron familias de acuerdo con los sexos de los niños. Los datos se recogieron en Alemania en el siglo XIX, cuando eran comunes las familias grandes. La Tabla 3.7.1 muestra los resultados de 6.115 familias con 12 niños²⁰.

Es interesante considerar si la variación observada entre las familias se puede explicar mediante el modelo de pruebas independientes. Exploraremos esta cuestión ajustando una distribución binomial a los datos.

El primer paso para ajustar la distribución binomial es determinar un valor de $p = \Pr\{\text{niño}\}$. Una posibilidad sería suponer que $p = 0,50$. Sin embargo, como se sabe que la razón de sexo de los humanos en el nacimiento no es exactamente 1:1 (de hecho, favorece ligeramente los niños), no haremos esa suposición. En vez de eso,

Tabla 3.7.1 Razón de sexos en 6.115 familias con 12 niños		
Número de		Frecuencia observada (número de familias)
Niños	Niñas	
0	12	3
1	11	24
2	10	104
3	9	286
4	8	670
5	7	1.033
6	6	1.343
7	5	1.112
8	4	829
9	3	478
10	2	181
11	1	45
12	0	7
		6.115

«ajustaremos» un valor de p a los datos. Es decir, determinaremos un valor de p que se ajuste mejor a los datos. Observamos que número total de hijos en todas las familias es

$$(12)(6.115) = 73.380 \text{ hijos}$$

Entre estos hijos, el número de niños es

$$(3)(0) + (24)(1) + \dots + (12)(7) = 38.100 \text{ hijos}$$

Por tanto, el valor de p que mejor se ajusta a los datos es

$$p = \frac{38.100}{73.380} = 0,519215$$

El siguiente paso es calcular las probabilidades de la fórmula de la distribución binomial con $n = 12$ y $p = 0,519215$. Por ejemplo, la probabilidad de tres niños y nueve niñas se calcula como

$${}_{12}C_3(p)^3(1-p)^9 = 220(0,519215)^3(0,480785)^9 \\ \approx 0,042269$$

Para comparar con los datos observados, convertiremos cada probabilidad en una frecuencia teórica o «esperada» multiplicándola por 6.115 (el número total de familias). Por ejemplo, el número esperado de familias con tres hijos y nueve hijas es

$$(6.115)(0,042269) \approx 258,5$$

Las frecuencias esperadas y observadas se muestran juntas en la Tabla 3.7.2. Dicha tabla muestra un ajuste razonable entre las frecuencias observadas y las predicciones de la distribución binomial. Pero un examen más detenido revela que las discrepancias, aunque no son grandes, siguen un patrón definido. Contienen más hermanos unisexuales, o preponderancia unisexual, que la que sería de esperar. De hecho, las frecuencias observadas son mayores que las frecuencias esperadas en nueve tipos de familias en las cuales predomina un sexo o el otro, mientras que las frecuencias observadas son menores que las esperadas en cuatro tipos de familias más «equilibradas». Este patrón se revela más claramente en la última columna de la Tabla 3.7.2, que muestra el signo de la diferencia entre la frecuen-

cia observada y la frecuencia esperada. Por tanto, la distribución observada de la proporción de sexos tiene «colas» más grandes y una «parte media» menor que la distribución binomial que mejor se ajusta.

Tabla 3.7.2 Razón de sexos y frecuencias binomiales esperadas				
Número de		Frecuencia observada	Frecuencia esperada	Signo de (obs. – esp.)
Niños	Niñas			
0	12	3	0,9	+
1	11	24	12,1	+
2	10	104	71,8	+
3	9	286	258,5	+
4	8	670	628,1	+
5	7	1.033	1.085,2	–
6	6	1.343	1.367,3	–
7	5	1.112	1.265,6	–
8	4	829	854,3	–
9	3	478	410,0	+
10	2	181	132,8	+
11	1	45	26,1	+
12	0	7	2,3	+
		6.115	6.115,0	

El patrón sistemático de desviaciones de la distribución binomial sugiere que la variación observada entre las familias no se puede explicar completamente mediante el modelo de pruebas independientes*. ¿Qué factores podrían dar cuenta de la discrepancia? Esta intrigante pregunta ha estimulado a investigadores a realizar análisis más detallados de estos datos. Presentaremos brevemente algunos de los aspectos.

Una explicación del exceso de familias predominantemente unisexuales es que la probabilidad de producir un niño puede variar entre las familias. Si p varía de una familia a otra, entonces el sexo aparentará «desplazarse» entre las familias en el sentido de que el número de familias predominantemente unisexuales aumentará. Para visualizar claramente este efecto, consideremos el conjunto de datos ficticios que se muestra en la Tabla 3.7.3.

En el conjunto de datos ficticios, hay $(3.175)(12) = 38.100$ niños entre los 73.380 hijos, lo mismo que en el conjunto de datos real. Por tanto, el valor de p que se ajusta mejor es el mismo ($p = 0.519215$) y las frecuencias binomiales esperadas son las mismas que en la Tabla 3.7.2. El conjunto de datos ficticios contiene solo familias con hijos de un solo sexo y por tanto es un ejemplo extremo de «desplazamiento» del sexo en familias. El conjunto de datos real muestra el mismo fenómeno de manera más débil. Una explicación del conjunto de datos ficticios sería que algunas familias tienen solamente niños ($p = 1$) y otras familias tienen solamente niñas ($p = 0$). De forma paralela, una explicación de los datos reales sería que p varía ligeramente entre las familias. La variación de p es verosímil biológicamente, incluso aunque el mecanismo que causa la variación no haya sido todavía descubierto.

Una explicación alternativa del mayor número de familias sexualmente homogéneas sería que los sexos de los hijos en una familia son literalmente dependientes entre sí, en el sentido de que la determinación del sexo de un hijo en particular está algo influenciado por los sexos de los hijos anteriores. Esta explicación no es verosímil en términos biológicos debido a que es difícil imaginar como el sistema biológico podría «recordar» los sexos de la descendencia previa. ■

* Una prueba de bondad de ajuste chi-cuadrado de la distribución binomial muestra que hay una fuerte evidencia de que las diferencias entre las frecuencias observadas y esperadas no sean debidas al error aleatorio en el proceso de muestreo. Explicaremos el concepto de pruebas de bondad de ajuste en el Capítulo 9.

Tabla 3.7.3 Datos ficticios de razón de sexos y frecuencias binomiales esperadas

Número de		Frecuencia observada	Frecuencia esperada	Signo de (obs. – esp.)
Niños	Niñas			
0	12	2.940	0,9	+
1	11	0	12,1	–
2	10	0	71,8	–
3	9	0	258,5	–
4	8	0	628,1	–
5	7	0	1.085,2	–
6	6	0	1.367,3	–
7	5	0	1.265,6	–
8	4	0	854,3	–
9	3	0	410,0	–
10	2	0	132,8	–
11	1	0	26,1	–
12	0	3.175	2,3	+
		6.115	6.115,0	

El Ejemplo 3.7.1 muestra que la pobreza de ajuste al modelo de pruebas independientes puede ser biológicamente interesante. Debemos resaltar, sin embargo, que la mayoría de las aplicaciones estadísticas de la distribución binomial proceden del supuesto de que el modelo de pruebas independientes es aplicable. En una aplicación típica, los datos se pueden ver como resultado de un *único* conjunto de n pruebas. Conjuntos de datos como los de la razón de sexo en las familias, que se refieren a *muchos* conjuntos de $n = 12$ pruebas, no se encuentran a menudo.

Ejercicios 3.7.1-3.7.3

3.7.1 Los datos que se presentan a continuación sobre familias con 6 hijos están tomados del mismo estudio de las familias con 12 hijos del Ejemplo 3.7.1. Ajuste una distribución binomial a los datos. (Redondee las frecuencias espe-

Número de		Número de familias
Niños	Niñas	
0	6	1.096
1	5	6.233
2	4	15.700
3	3	22.221
4	2	17.332
5	1	7.908
6	0	1.579
		72.069

radas a un dígito decimal). Compare con los resultados del Ejemplo 3.7.1. ¿Qué características comparten los dos conjuntos de datos?

3.7.2 Un método importante para estudiar las sustancias que causan mutaciones consiste en matar a ratones hembra 17 días después de aparearse y examinar sus úteros en busca de embriones vivos y muertos. El método clásico de análisis de estos datos suponen que la supervivencia o muerte de cada embrión constituye una prueba independiente binomial. La tabla que sigue, extraída de un estudio más amplio, proporciona datos de 310 hembras, y en todas ellas sus úteros contenían nueve embriones. Todos los animales fueron tratados del mismo modo (como controles)²¹.

- Ajuste una distribución binomial a los datos observados. (Redondee las frecuencias esperadas a un dígito decimal).
- Interprete la relación entre las frecuencias esperadas y observadas. ¿Arrojan los datos sospechas sobre la suposición clásica?

Número de embriones		Número de ratones hembra
Muertos	Vivos	
0	9	136
1	8	103
2	7	50
3	6	13
4	5	6
5	4	1
6	3	1
7	2	0
8	1	0
9	0	0
		310

3.7.3 Los estudiantes de una clase grande de botánica realizaron un experimento sobre la germinación de semillas del cactus Saguaro. Como parte del experimento, cada estudiante plantó cinco semillas en un pequeño tiesto, puso el tiesto cerca de una ventana y comprobó cada día la germinación (aparición de brotes). Los resultados de la clase al séptimo día después de plantar fueron los que se muestran en la tabla²².

- (a) Ajuste una distribución binomial a los datos. (Redondee las frecuencias esperadas a un dígito decimal).
- (b) Se habló con dos estudiantes, Fran y Bob, antes de la clase. Todas las semillas de Fran habían germinado el

Número de semillas		Número de estudiantes
Germinadas	No germinadas	
0	5	17
1	4	53
2	3	94
3	2	79
4	1	33
5	0	4
		280

séptimo día, mientras que ninguna de las de Bob había germinado. Bob se preguntaba si había hecho algo incorrecto. Con la perspectiva ganada al ver los resultados de los 280 estudiantes, ¿qué le diría a Bob? (*Sugerencia:* ¿puede explicarse la variación entre estudiantes mediante la hipótesis de que algunas de las semillas eran buenas y algunas malas, y cada estudiante recibe cinco semillas aleatoriamente escogidas?)

- (c) Invente un conjunto de datos ficticio de 280 estudiantes, con el mismo porcentaje global de germinación de los datos observados dados en la tabla, pero en el que todos los estudiantes obtienen o bien los resultados de Fran (perfectos) o los resultados de Bob (nada). ¿Cómo respondería a la diferencia de Bob si los datos reales fueran como este conjunto de datos ficticio?

Ejercicios suplementarios 3.S.1-3.S.10

3.S.1 En Estados Unidos, el 10 % de las chicas adolescentes tienen deficiencia de hierro²³. Suponga que se escogen aleatoriamente dos chicas adolescentes. Calcule la probabilidad de que

- (a) Ambas chicas tengan deficiencia de hierro.
- (b) Una de las chicas tenga deficiencia de hierro y la otra no.

3.S.2 En la preparación de un estudio ecológico sobre ciempiés, se dividió el suelo de un bosque de hayas en un gran número de cuadrados de un pie de lado²⁴. En un cierto momento, la distribución de ciempiés en los cuadrados es como se muestra en la tabla.

Suponga que se escoge aleatoriamente un cuadrado, y sea Y el número de ciempiés en el cuadrado elegido. Calcule

- (a) $\Pr\{Y = 1\}$.
- (b) $\Pr\{Y \geq 2\}$.

Número de ciempiés	Frecuencia porcentual (% de cuadrados)
0	45
1	36
2	14
3	4
4	1
	100

3.S.3 En relación con la distribución de ciempiés dada en el Ejercicio 3.S.2, suponga que se escogen aleatoriamente cinco cuadrados. Calcule la probabilidad de que tres de los cuadrados contengan ciempiés y dos no.

3.S.4 En relación con la distribución de ciempiés dada en el Ejercicio 3.S.2, suponga que se escogen aleatoriamente cinco cuadrados. Calcule el valor esperado (es decir, la media) del número de cuadrados que contienen al menos un ciempiés.

3.S.5 El pelo ondulado en los ratones es un rasgo genético recesivo. Si se aparean ratones con pelo ondulado con ratones con pelo liso (heterocigóticos), cada descendiente tiene una probabilidad de $\frac{1}{2}$ de tener pelo ondulado²⁵. Considere un gran número de apareamientos en los que se producen camadas de cinco descendientes. Indique en qué porcentaje de cada camada

- (a) Hay dos ratones de pelo ondulado y tres de pelo liso.
- (b) Tres o más ratones son de pelo liso.
- (c) Son todos del mismo tipo (o todos de pelo ondulado o todos de pelo liso).

3.S.6 Un cierto medicamento causa daños en el riñón en el 1 % de los pacientes. Suponga que el medicamento se prueba con 50 pacientes. Calcule la probabilidad de que

- (a) Ninguno de los pacientes experimente daños en el riñón.
 - (b) Uno o más pacientes experimente daños en el riñón.
- [Sugerencia: use el apartado (a) para responder al apartado (b)].

3.S.7 Refiriéndonos al Ejercicio 3.S.6, suponga ahora que el medicamento se va a probar en n pacientes, y sea E el suceso de que ocurren daños en el riñón en uno o más de los pacientes. La probabilidad $\Pr\{E\}$ es útil para establecer criterios de seguridad de medicamentos.

- (a) Calcule $\Pr\{E\}$ para $n = 100$.
- (b) ¿Qué valor debe tener n para que $\Pr\{E\}$ no sea mayor que 0,95?

3.S.8 Para estudiar la capacidad de la gente para engañar a los detectores de mentiras, los investigadores utilizan a veces la técnica del «conocimiento culpable»²⁶. Ciertos sujetos

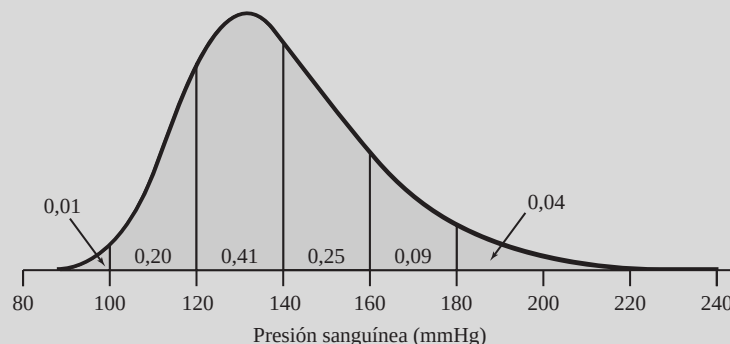
memorizan seis palabras comunes. Otros sujetos no memorizan ninguna palabra. Cada sujeto pasa después la prueba del polígrafo (el detector de mentiras) como sigue. El experimentado lee, en orden aleatorio, 24 palabras: las seis palabras «críticas» (la lista memorizada) y, para cada palabra crítica, tres palabras «de control» con significados similares o relacionados. Si el sujeto ha memorizado las seis palabras, éste intenta ocultar ese hecho. El sujeto se indica como «fallo» en una palabra crítica si su respuesta electrodérmica es mayor en la palabra crítica que en cualquiera de las tres palabras de control. Por tanto, en cada una de las seis palabras críticas, incluso un sujeto inocente tendría una probabilidad de fallo del 25 %. Suponga que un sujeto se etiqueta como «culpable» si falla en cuatro o más de las seis palabras críticas. Si se realizara la prueba sobre un sujeto inocente, ¿cuál es la probabilidad de que él o ella sean etiquetados como «culpable»?

3.S.9 La curva de densidad que se muestra a continuación representa la distribución de las presiones sanguíneas sistólicas en una población de hombres de mediana edad²⁷. Las áreas bajo la curva se muestran en la figura. Suponga que se selecciona aleatoriamente un hombre de la población, y sea Y su presión sanguínea. Calcule

- (a) $\Pr\{120 < Y < 160\}$.
- (b) $\Pr\{Y < 120\}$.
- (c) $\Pr\{Y > 140\}$.

3.S.10 En relación con la distribución de la presión sanguínea del Ejercicio 3.S.9, suponga que se seleccionan aleatoriamente cuatro hombres de la población. Calcule la probabilidad de que

- (a) Los cuatro tengan presiones sanguíneas mayores que 140 mmHg.
- (b) Tres de ellos tengan presiones sanguíneas mayores que 140, y uno tenga una presión sanguínea de 140 o menos.



Notas

1. Basado en un artículo del Neonatal Inhaled Nitric Oxide Study Group (1997). Véase Inhaled nitric oxide in full-term and nearly full-term infants with hypoxic respiratory failure. *New England Journal of Medicine* **336**, 597-604.
2. Población ficticia pero realista. Adaptado de Hubbs, C. L. y Schultz, L. P. (1932). *Cottus tubulatus*, a new sculpin from Idaho. *Occasional Papers of the Museum of Zoology, University of Michigan* **242**, 1-9. Datos reproducidos en Simpson, G. G., Roe, A. y Lewontin, R. C. (1960). *Quantitative Zoology*. New York: Harcourt, Brace. p. 81.
3. www.bloodbook.com/world-abo.html
4. Esta tabla es una versión modificada de los datos adaptados de Ammon, O. (1899). *Zur Anthropologie der Badener*. Jena: G. Fischer. Los datos de Ammon aparecen en Goodman, L. A. y Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association* **49**, 732-764. Los números de la tabla se han redondeado como ayuda en la exposición.
5. Datos no publicados cortesía de Diana Zumas y Lisa Yasuhara, Oberlin College.
6. Adaptado de Taira, D. A., Safran, D. G., Seto, T. B., Rogers, W. H. y Tarlov, A. R. (1997). The relationship between patient income and physician discussion of health risk behaviors. *Journal of the American Medical Association* **278**, 1412-1417.
7. La población es ficticia pero se parece a la población de mujeres americanas de edades entre 18 y 24, excluyendo las diabéticas o sospechosas de serlo, como se indica en Gordon, T. (1964). Glucose tolerance of adults, United States 1960-62. *U.S. National Center for Health Statistics, Vital and Health Statistics*, serie 11, n.º 2. Washington, D.C.: U.S. Department of Health, Education and Welfare.
8. Meyer, W. H. (1930). Diameter distribution series in even-aged forest stands. *Yale University School of Forestry Bulletin* **28**. La curva se ajusta en Bliss, C. I. y Reinker, K. A. (1964). A lognormal approach to diameter distributions in even-aged stands. *Forest Science* **10**, 350-360.
9. Pearson, K. (1914). On the probability that two independent distributions of frequency are really samples of the same population, with reference to recent work on the identity of trypanosome strains. *Biometrika* **10**, 85-143. Reimpreso con permiso de the Biometrika Trustees.
10. Adaptado de datos no publicados cortesía de Gloria Zender, Oberlin College.
11. Situación ficticia pero realista. Basado en datos dados por Lack, D. (1948). Natural selection and family size in the starling. *Evolution* **2**, 95-110. Datos reproducidos por Riclefs, R. E. (1973). *Ecology*. Newton, Mass.: Chiron Press, p. 37.
12. Adaptado de datos no publicados cortesía de Marni Hansill, Oberlin College.
13. Halpine, T. y Kerr, S. J. (1986). Mutant allele frequencies in the cat population of Omaha, Nebraska. *The Journal of Heredity* **77**, 460-462.
14. Este es uno de los cruces realizados por Gregor Mendel en sus estudios clásicos sobre la herencia; plantas heterocigóticas (de semilla amarilla porque el amarillo es dominante) se cruzan entre sí.
15. Datos ficticios pero realistas. Véase Hutchison, J. G. P., Johnston, N. M., Plevey, M. V. P., Thangkhiew, I. y Aidney, C. (1975). Clinical trial of Mebendazole, a broad-spectrum anthelmintic. *British Medical Journal* **2**, 309-310.
16. Población ficticia pero realista. Adaptado de Owen, D. F. (1963). Polymorphism and population density in the African land snail, *Limicolaria martensiana*. *Science* **140**, 666-667.
17. Mathews, T. J. y Hamilton, B. E. (2005). Trend analysis of the sex ratio at birth in the United States. *National Vital Statistics Reports* **53**, n.º 20. Hyatts ville Md.; National Center for Health Statistics. La razón de sexos varía ligeramente con el tiempo y la raza.
18. Adaptado de una presentación en Galen, R. S. y Gambino, S. R. (1980). *Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses*. New York: Wiley, pp. 71-74.
19. Esto sería cierto para algunas poblaciones de ciudades centrales. Véase Annett, J. L., Mahaffey, K. R., Cox, D. H. y Roberts, J. (1982). Blood lead levels for persons 6 months-74 years of age: United States, 1976-80. *U.S. National Center for Health Statistics, Advance Data from Vital and Health Statistics*, n.º 79. Hyattsville, Md.; U.S. Department of Health and Human Services.
20. Geissler, A. (1889). Beitrage zur Frage des Geschlechtsverhältnisses der Geborenen. *Zeitschrift des K. Sächsischen Statistischen Bureaus* **35**, 1-24. Datos reproducidos por Edwards, A. W. F. (1958). An analysis of Geissler's data on the human sex ratio. *Annals of Human Genetics* **23**, 6-15. Los datos se comentan también en Stern, C. (1960). *Human Genetics*. San Francisco: Freeman.
21. Haseman, J. K. y Soares, E. R. (1976). The distribution of fetal death in control mice and its implications on statistical tests for dominant lethal effects. *Mutation Research* **41**, 277-288.

22. Datos cortesía de S. N. Postlethwaite.
23. Adaptado de Looker, A., *et al.* (1997). Prevalence of iron deficiency in the United States. *Journal of the American Medical Association* 277, 973-976.
24. Situación ficticia pero realista. Véase Krebs, C. J. (1972). *Ecology: The Experimental Analysis of Distribution and Abundance*. New York: Harper & Row, p. 142.
25. Véase Mather, K. (1943). *Statistical Analysis in Biology*. London: Methuen, p. 38.
26. La técnica se describe en Waid, W. M., Orne, E. C., Cook, M. R. y Orne, M. T. (1981). Meprobamate reduces accuracy of physiological detection of deception. *Science* 212, 71-73.
27. Población ficticia pero realista, que se parece mucho a la población de hombres entre 45 y 59 años descrita en Roberts, J. (1975). Blood pressure of persons 18-74 years, United States, 1971-72. *U.S. National Center for Health Statistics, Vital and Health Statistics*, serie 11, n.º 150. Washington, D.C.: U.S. Department of Health, Education and Welfare.

LA DISTRIBUCIÓN NORMAL

Objetivos

En este capítulo estudiaremos la distribución normal, incluyendo:

- el uso de la curva normal para modelar distribuciones;
- el cálculo de probabilidades utilizando la curva normal;
- evaluación de la normalidad de conjuntos de datos mediante el uso de gráficas de probabilidad normales.

4.1 Introducción

En el Capítulo 2 presentamos la idea de considerar un conjunto de datos como una muestra de una población. En la Sección 3.4 vimos que la distribución poblacional de una variable cuantitativa Y se puede describir mediante su media μ y su desviación típica σ y también mediante una curva de densidad, que representa las frecuencias relativas como áreas bajo la curva. En este capítulo estudiaremos el tipo más importante de curvas de densidad: la **curva normal**. La curva normal es una curva simétrica con «forma de campana», cuya forma exacta describiremos posteriormente. Una distribución representada por una curva normal se denomina **distribución normal**.

La familia de las distribuciones normales tiene dos papeles en aplicaciones estadísticas. Su uso directo es como una aproximación adecuada de la distribución de una variable observada Y . El segundo papel de la distribución normal es más teórico y se explorará en el Capítulo 5.

Se presenta a continuación un ejemplo de una distribución de una población natural que se puede aproximar por una distribución normal.

Ejemplo 4.1.1 Colesterol en el suero

La relación entre la concentración de colesterol en la sangre y la aparición de enfermedad del corazón ha sido objeto de muchas investigaciones. Como parte de un estudio gubernamental sobre la salud, unos investigadores midieron los niveles de colesterol en el suero de una muestra grande de americanos, incluyendo niños. La distribución para los niños entre 12 y 14 años de edad se puede aproximar bastante bien mediante una curva normal de media $\mu = 162$ mg/dl y desviación típica $\sigma = 28$ mg/dl. La Figura 4.1.1 muestra un histograma basado en una muestra de 727 niños entre 12 y 14 años de edad, con la curva normal superpuesta¹. ■

Para indicar cómo se relaciona la media μ y la desviación típica σ con la curva normal, la Figura 4.1.2 muestra la curva normal de la distribución de colesterol en el suero del Ejemplo 4.1.1, con marcas a la distancia de 1, 2 y 3 desviaciones típicas de la media.

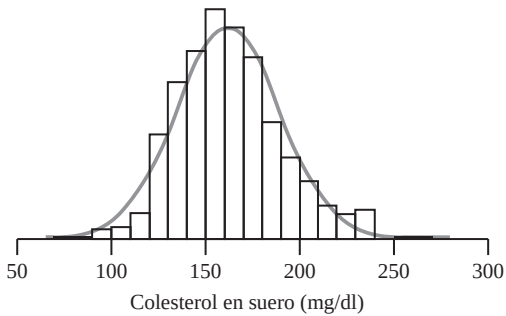


Figura 4.1.1 Distribución del colesterol en suero en 727 niños entre 12 y 14 años de edad

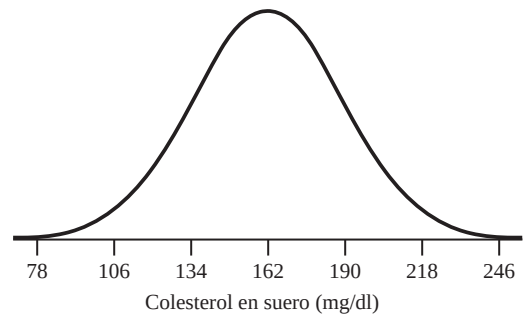


Figura 4.1.2 Distribución del colesterol en suero con $\mu = 162$ mg/dl y $\sigma = 28$ mg/dl

La curva normal se puede usar para describir la distribución de una variable observada Y de dos formas: (1) como una aproximación suave a un histograma basado en una muestra de valores de Y y (2) como una representación idealizada de la distribución poblacional de Y . Las curvas normales de las Figuras 4.1.1 y 4.1.2 se pueden interpretar de cualquiera de las dos formas. Por simplicidad, en el resto de este capítulo consideraremos la curva normal como la representación de la distribución poblacional.

Más ejemplos

Presentaremos ahora tres ejemplos más de curvas normales que describen aproximadamente poblaciones reales. En cada figura, la escala del eje horizontal tiene marcas situadas en la media y en valores separados una desviación típica.

Ejemplo 4.1.2 Espesor de cáscaras de huevo

En la producción comercial de huevos, la rotura es un problema importante. Por tanto, el espesor de las cáscaras de huevos es una variable importante. En un estudio se observó que el espesor de las cáscaras de huevos producidos por un gran número de gallinas White Leghorn seguía aproximadamente una distribución normal de media $\mu = 0,38$ mm y desviación típica $\sigma = 0,03$ mm. La distribución se muestra en la Figura 4.1.3². ■

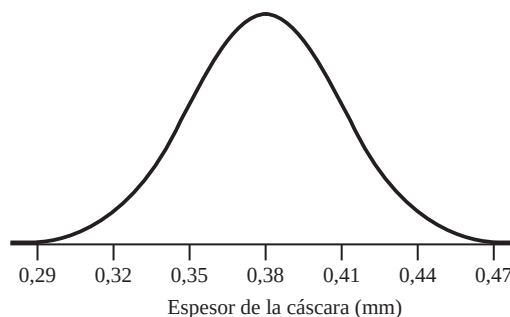


Figura 4.1.3 Distribución normal del espesor de las cáscaras de huevos, con $\mu = 0,38$ mm y $\sigma = 0,03$ mm

Ejemplo 4.1.3 Tiempos entre disparos de células nerviosas

En ciertas células nerviosas se observó que las descargas eléctricas espontáneas se repetían de forma tan rítmica que se denominaban «picos de reloj». Los instantes de esos picos, incluso siendo remarcablemente regulares, presentaban variación. En un estudio, se observó que los intervalos entre picos (en milisegundos) de una única mosca doméstica (*Musca domestica*) seguían aproximadamente una distribución normal de media $\mu = 15,6$ ms y desviación típica $\sigma = 0,4$ ms. Esta distribución se muestra en la Figura 4.1.4³. ■

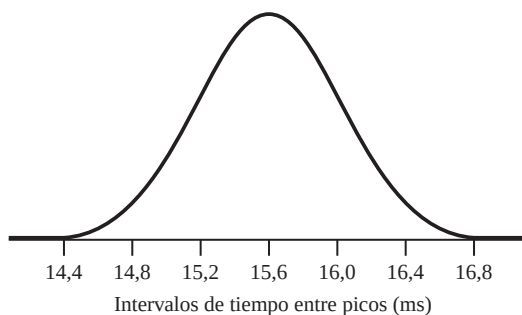


Figura 4.1.4 Distribución normal de los intervalos de tiempo entre picos, con $\mu = 15,6$ ms y $\sigma = 0,4$ ms

Los ejemplos anteriores han ilustrado tipos muy diferentes de poblaciones. En el Ejemplo 4.1.3, la población completa consiste en medidas de una sola mosca. Otro tipo de población se denomina población de *error de medida*, que consiste en medidas repetidas de exactamente la misma cantidad. La desviación de cada medida del valor «correcto» se denomina error de medida. El error de medida no es el resultado de un error, sino que más bien es debido a la falta de precisión perfecta en el proceso de medida o en los instrumentos de medida. Las distribuciones de los errores de medida a menudo son aproximadamente normales. En este caso la media de la distribución de las medidas repetidas de la misma cantidad es el valor verdadero de dicha cantidad (suponiendo que el instrumento de medida está correctamente calibrado), y la desviación típica de la distribución indica la precisión del instrumento. En el Ejemplo 2.2.12 se presentó una distribución de error de medida. Sigue a continuación otro ejemplo.

Ejemplo 4.1.4 Error de medida

Cuando se utiliza un cierto instrumento electrónico para contar partículas como glóbulos blancos, la distribución del error de medida es aproximadamente normal. En el caso de los glóbulos blancos, la desviación típica de las medidas repetidas basadas en el mismo espécimen de sangre es aproximadamente el 1,4 % del valor verdadero. Por tanto, si el valor verdadero de un cierto espécimen de sangre era de 7.000 células/mm³, entonces la desviación típica sería de aproximadamente 100 células/mm³ y la distribución de las medidas repetidas del espécimen sería semejante a la Figura 4.1.5⁴. ■

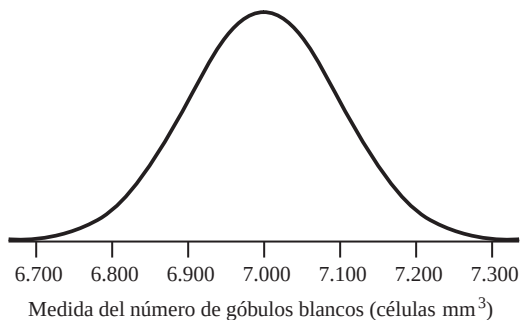


Figura 4.1.5 Distribución normal de medidas repetidas de número de glóbulos blancos en un espécimen de sangre cuyo verdadero valor es $\mu = 7.000$ células/mm³ y $\sigma = 100$ células/mm³

4.2 Las curvas normales

Como muestran los ejemplos de la Sección 4.1, hay muchas curvas normales. Cada curva normal concreta se caracteriza por su media y su desviación típica. Si la variable Y sigue una distribución normal de media μ y desviación típica σ , entonces es habitual escribir $Y \sim N(\mu, \sigma)$. Todas las curvas normales se pueden expresar mediante una sola fórmula.

Aunque no utilizaremos directamente la fórmula en este libro, la presentaremos aquí, tanto como un tema de interés como para recalcar que una curva normal no es cualquier curva simétrica, sino un tipo *específico* de curva simétrica.

Si una variable Y sigue una distribución normal de media μ y desviación típica σ , entonces la curva de densidad de la distribución de Y está dada por la siguiente fórmula:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

Esta función, $f(y)$, se denomina *función de densidad* de la distribución y expresa la altura de la curva en función de la posición y en el eje y . Las cantidades e y π que aparecen en la fórmula son constantes, con e aproximadamente igual a 2,71 y π aproximadamente igual a 3,14.

La Figura 4.2.1 muestra una gráfica de una curva normal. La forma de la curva es como la de una campana simétrica, centrada en $y = \mu$. La dirección de curvatura es hacia abajo (como un cuenco invertido) en la parte central de la curva, y hacia arriba en la parte de las colas. Los puntos de inflexión (es decir, donde la curvatura cambia de dirección) son $y = \mu - \sigma$ e $y = \mu + \sigma$. Nótese que la curva es casi lineal en los alrededores de esos puntos. En principio la curva se extiende hasta $+\infty$ y $-\infty$, y nunca alcanza realmente el eje y . Sin embargo, la altura de la curva es muy pequeña para valores de y alejados más de tres desviaciones típicas de la media. El área bajo la curva es exactamente igual a 1. (Nota: Pede parecer paradójico que una curva pueda encerrar un área finita, incluso aunque nunca descienda a tocar el eje y . Esta paradoja aparente se clarifica en el Apéndice 4.1).

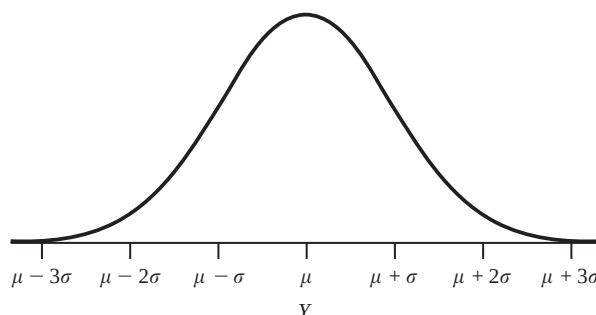


Figura 4.2.1 Una curva normal de media μ y desviación típica σ

Todas las curvas normales tienen la misma forma esencial, en el sentido de que se pueden hacer idénticas mediante una selección adecuada de las escalas vertical y horizontal de cada una. (Por ejemplo, nótese que las curvas de las Figuras 4.1.4-4.1.5 parecen idénticas). Pero curvas normales con diferentes valores de μ y σ no parecerán idénticas si se dibujan todas en la misma escala, como se ilustra en la Figura 4.2.2. La posición de la curva normal en el eje y está gobernada por μ , ya que la curva está centrada en $y = \mu$. La anchura de la curva está gobernada por σ . La altura de la curva también está determinada por σ . Como el área bajo cada curva debe ser igual a 1, una curva con un valor más pequeño de σ debe tener una altura mayor. Esto refleja el hecho de que los valores de Y están más altamente concentrados cerca de la media cuando la desviación típica es pequeña.

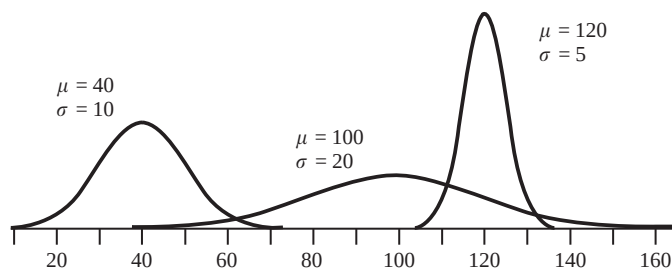


Figura 4.2.2 Tres curvas normales con diferentes medias y desviaciones típicas

4.3 Áreas bajo una curva normal

Como se explicó en la Sección 3.4, una curva de densidad se puede interpretar cuantitativamente en términos de áreas bajo la curva. Aunque las áreas se pueden estimar aproximadamente a ojo, en algunos casos es deseable tener una información bastante precisa sobre dichas áreas.

La escala tipificada

Las áreas bajo una curva normal se han calculado matemáticamente y están tabuladas aquí para su uso práctico. El uso de esta información tabulada se simplifica mucho por el hecho de que todas las curvas normales se pueden hacer equivalentes, con respecto a las áreas bajo ellas, mediante un cambio de escala adecuado del eje horizontal. La variable con la escala cambiada se denomina Z . La relación entre las escalas se muestra en la Figura 4.3.1.

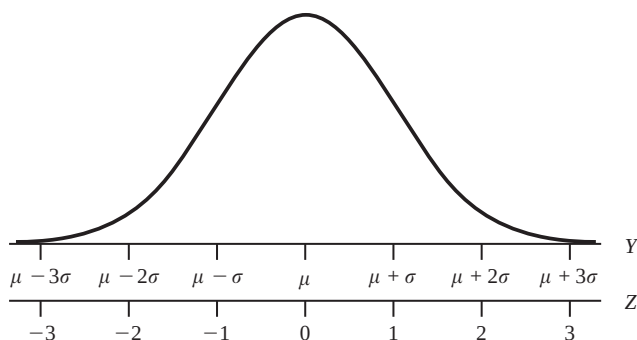


Figura 4.3.1 Una curva normal que muestra la relación entre la escala natural (Y) y la escala tipificada (Z)

Como indica la Figura 4.3.1, la escala Z mide desviaciones típicas desde la media: $z = 1,0$ corresponde a 1,0 desviación típica por encima de la media; $z = -2,5$ corresponde a 2,5 desviaciones típicas por debajo de la media, y así sucesivamente. La escala Z se denomina **escala tipificada**.

La correspondencia entre la escala Z y la escala Y se puede expresar mediante la fórmula que se presenta en el recuadro siguiente.

Fórmula de tipificación

$$Z = \frac{Y - \mu}{\sigma}$$

La variable Z se denomina **normal tipificada** o **normal estándar** y su distribución sigue una curva normal de media cero y desviación típica uno. La Tabla 3 al final del libro proporciona las áreas bajo la curva normal tipificada, con distancias en el eje horizontal medidas en la escala Z . Cada área tabulada en la Tabla 3 es el área bajo la curva normal tipificada por debajo de un valor concreto de z . Por ejemplo, para $z = 1,53$, el área tabulada es 0,9370. Esta área se muestra sombreada en la Figura 4.3.2.

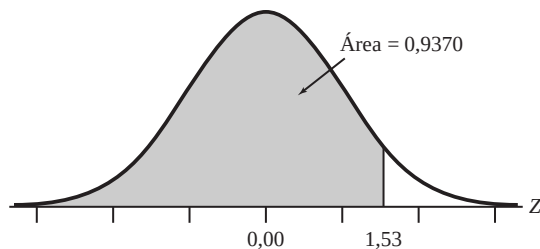


Figura 4.3.2 Ilustración del uso de la Tabla 3

Si deseamos obtener el área por encima de un valor dado de z , restaremos el área tabulada de 1. Por ejemplo, el área por encima de $z = 1,53$ es $1,0000 - 0,9370 = 0,0630$ (Figura 4.3.3).

Para obtener el área entre dos valores de z (también comúnmente denominadas **puntuaciones z**), se restan las áreas dadas en la Tabla 3. Por ejemplo, para obtener el área bajo la curva Z entre $z = -1,2$ y $z = 0,8$ (Figura 4.3.4), tomamos el área por debajo de $0,8$, que es $0,7881$, y restamos el área debajo de $-1,2$, que es $0,1151$, obteniendo $0,7881 - 0,1151 = 0,6730$.

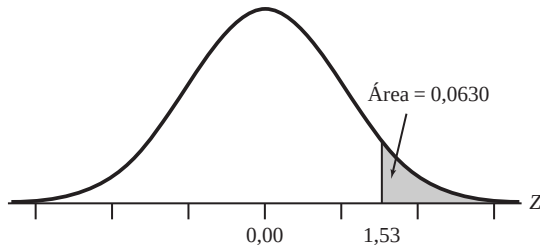


Figura 4.3.3 Área bajo una curva normal tipificada por encima de $1,53$

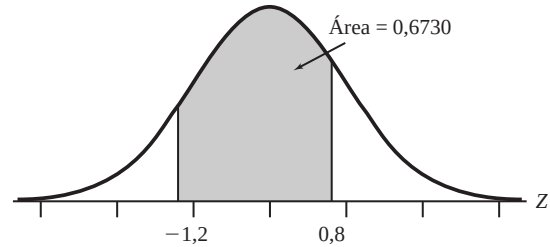


Figura 4.3.4 Área bajo una curva normal tipificada entre $-1,2$ y $0,8$

Utilizando la Tabla 3, vemos que el área bajo la curva normal entre $z = -1$ y $z = +1$ es $0,8413 - 0,1578 = 0,6826$. Por tanto, para cualquier distribución normal, aproximadamente el 68 % de las observaciones están dentro de ± 1 desviación típica alrededor de la media. Asimismo, el área bajo la curva normal entre $z = -2$ y $z = +2$ es $0,9772 - 0,0228 = 0,9544$ y el área bajo la curva normal entre $z = -3$ y $z = +3$ es $0,9987 - 0,0013 = 0,9974$. Esto significa que en cualquier distribución normal aproximadamente el 95 % de las observaciones están dentro de ± 2 desviaciones típicas alrededor de la media y aproximadamente el 99,7 % de las observaciones están dentro de ± 3 desviaciones típicas alrededor de la media (véase la Figura 4.3.5). Por ejemplo, aproximadamente el 68 % de los valores de colesterol en suero de la distribución idealizada de la Figura 4.1.2 están entre 134 mg/dl y 190 mg/dl, aproximadamente el 95 % están entre 106 mg/dl y 218 mg/dl y virtualmente todos están entre 78 mg/dl y 246 mg/dl. La Figura 4.3.6 muestra estos porcentajes.

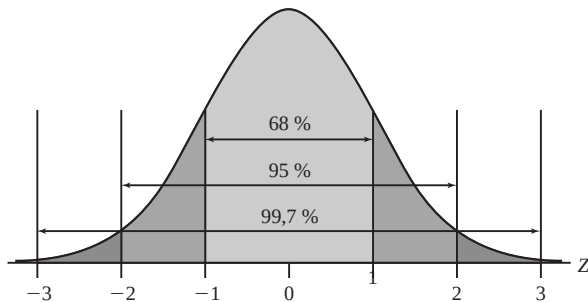


Figura 4.3.5 Áreas bajo una curva normal entre -1 y $+1$, entre -2 y $+2$ y entre -3 y $+3$

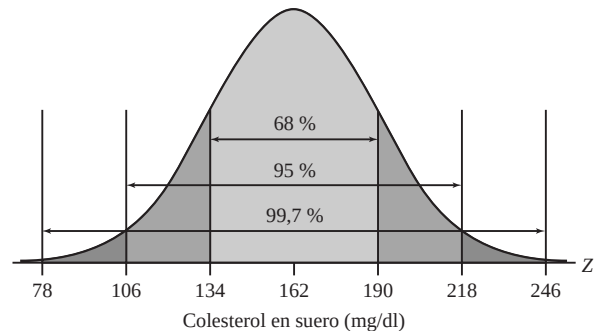


Figura 4.3.6 La regla 68/95/99,7 de la distribución de colesterol en suero

Si la variable Y sigue una distribución normal, entonces:

aproximadamente el 68 % de los valores de y están dentro de un intervalo de ± 1 DT alrededor de la media.

aproximadamente el 95 % de los valores de y están dentro de un intervalo de ± 2 DT alrededor de la media.

aproximadamente el 99,7 % de los valores de y están dentro de un intervalo de ± 3 DT alrededor de la media.

Estas afirmaciones proporcionan una interpretación muy concreta de la desviación típica en casos donde una distribución es aproximadamente normal. (De hecho, las afirmaciones son a menudo aproximadamente ciertas para distribuciones que son moderadamente no normales; esto es porque, en la Sección 2.6, estos porcentajes del 68 %, 95 % y >99 % se justificaron como «típicos» para distribuciones con «forma buena»).

Determinación de áreas de una curva normal

Aprovechando la escala tipificada, podemos utilizar la Tabla 3 para responder a preguntas detalladas sobre cualquier población normal cuando se especifica la media poblacional y la desviación típica poblacional. El ejemplo siguiente ilustra el uso de la Tabla 3. (Por supuesto, la población descrita en el ejemplo es un caso idealizado, ya que ninguna población real sigue *exactamente* una distribución normal).

Ejemplo 4.3.1 Longitudes de peces

En una cierta población de arenques *Pomolobus aestivalis*, las longitudes de los peces individuales siguen una distribución normal. La longitud media del pez es 54,0 mm, y la desviación típica es 4,5 mm⁵. Utilizaremos la Tabla 3 para contestar a varias preguntas sobre la población.

- (a) ¿Qué porcentaje de los peces miden menos de 60 mm de longitud?

La Figura 4.3.7 muestra la curva de densidad de la población, con el área deseada sombreada. Para utilizar la Tabla 3, convertimos los límites del área de la escala Y a la escala Z como sigue:

Para $y = 60$, la puntuación z es

$$z = \frac{y - \mu}{\sigma} = \frac{60 - 54}{4,5} = 1,33$$

Por tanto, la pregunta «¿Qué porcentaje de los peces miden menos de 60 mm de longitud?» es equivalente a la pregunta «¿Cuál es el área bajo la curva normal tipificada por debajo del valor de z de 1,33?». Buscando $z = 1,33$ en la Tabla 3, encontramos que el área es 0,9082. Por tanto, el 90,82 % de los peces miden menos de 60 mm de longitud.

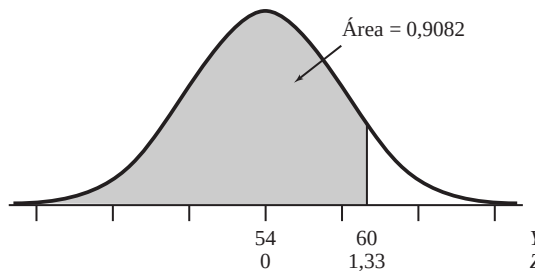


Figura 4.3.7 Área bajo una curva normal en el Ejemplo 4.3.1(a)

- (b) ¿Qué porcentaje de los peces miden más de 51 mm de longitud?

El valor tipificado de $y = 51$ es

$$z = \frac{y - \mu}{\sigma} = \frac{51 - 54}{4,5} = -0,67$$

Por tanto, la pregunta «¿Qué porcentaje de los peces miden más de 51 mm de longitud?» es equivalente a la pregunta «¿Cuál es el área bajo la curva normal tipificada por encima del valor de z de $-0,67$?». La Figura 4.3.8 muestra esta relación. Buscando $z = -0,67$ en la Tabla 3, encontramos que el área por debajo de $z = -0,67$ es 0,2514. Esta media del área por encima de $z = -0,67$ es $1 - 0,2514 = 0,7486$. Por tanto, el 74,86 % de los peces miden más de 51 mm de longitud.

- (c) ¿Qué porcentaje de los peces miden entre 51 y 60 mm de longitud?

La Figura 4.3.9 muestra el área deseada. Esta área se puede expresar como diferencia de 2 áreas que se obtienen de la Tabla 3. El área por debajo de $y = 60$ es de 0,9082, como se obtuvo en el apartado (a), y el área por debajo de $y = 51$ es 0,2514, como se obtuvo en el apartado (b). Entonces, el área deseada se calcula como

$$0,9082 - 0,2514 = 0,6568$$

Por tanto, el 65,68 % de los peces miden entre 51 y 60 mm de longitud.

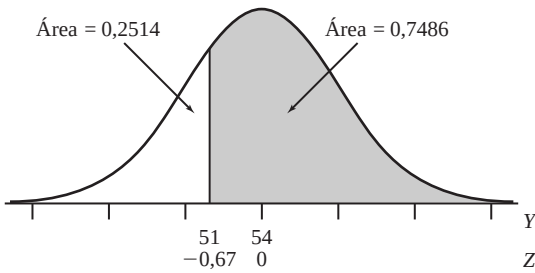


Figura 4.3.8 Área bajo una curva normal en el Ejemplo 4.3.1(b)

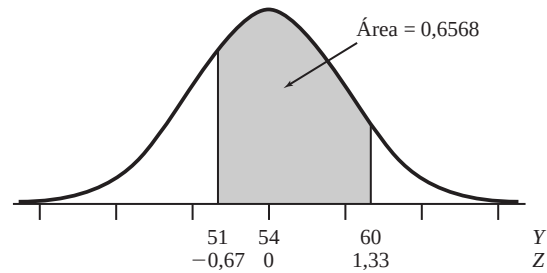


Figura 4.3.9 Área bajo una curva normal en el Ejemplo 4.3.1(c)

(d) ¿Qué porcentaje de los peces miden entre 58 y 60 mm de longitud?

La Figura 4.3.10 muestra el área deseada. Este área se puede expresar como diferencia de 2 áreas que se obtienen de la Tabla 3. El área por debajo de $y = 60$ es de 0,9082, como se obtuvo en el apartado (a). Para obtener el área por debajo de $y = 58$, calculamos primero el valor de z que corresponde a $y = 58$:

$$z = \frac{y - \mu}{\sigma} = \frac{58 - 54}{4,5} = 0,89$$

El área bajo la curva Z por debajo de $z = 0,89$ es 0,8133. Entonces, el área deseada se calcula como

$$0,9082 - 0,8133 = 0,0949$$

Por tanto, el 9,49 % de los peces miden entre 58 y 60 mm de longitud. ■

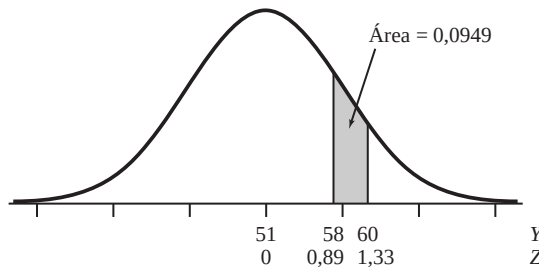


Figura 4.3.10 Área bajo una curva normal en el Ejemplo 4.3.1(d)

Cada uno de los porcentajes obtenidos en el Ejemplo 4.3.1 se puede interpretar también en términos de probabilidad. Sea la variable aleatoria Y que representa la longitud de un pez extraído aleatoriamente de la población. Entonces los resultados del Ejemplo 4.3.1 implican que

$$\Pr\{Y < 60\} = 0,9082$$

$$\Pr\{Y > 51\} = 0,7486$$

$$\Pr\{51 < Y < 60\} = 0,6568$$

y

$$\Pr\{58 < Y < 60\} = 0,0949$$

Entonces, la distribución normal se puede interpretar como una distribución de probabilidad continua.

Nótese que como la distribución normal idealizada es perfectamente continua, probabilidades como

$$\Pr\{Y > 48\} \quad \text{y} \quad \Pr\{Y \geq 48\}$$

son iguales (véase la Sección 3.4). Es decir,

$$\begin{aligned}\Pr\{Y \geq 48\} &= \Pr\{Y > 48\} + \Pr\{Y = 48\} \\ &= \Pr\{Y > 48\} + 0 \text{ (ya que } Y \text{ se considera continua)} \\ &= \Pr\{Y > 48\}\end{aligned}$$

Sin embargo, si las longitudes se midieran solo con una precisión de milímetros, entonces la variable medida sería realmente discreta, de forma que $\Pr\{Y > 48\}$ y $\Pr\{Y \geq 48\}$ diferirían algo entre sí. En casos donde esta discrepancia es importante, el cálculo se puede refinar teniendo en cuenta la discontinuidad de la distribución medida (veremos más tarde un ejemplo en la Sección 5.4).

Lectura inversa de la Tabla 3

Al determinar hechos sobre la distribución normal, algunas veces es necesario leer la Tabla 3 de forma «inversa», es decir, obtener el valor de z correspondiente al área dada en vez de la otra forma. Por ejemplo, supongamos que deseamos obtener el valor en la escala Z que deja por encima del 2,5 % de la distribución. Este número es 1,96, como se muestra en la Figura 4.3.11.

Nos resultará útil, para futuras referencias, introducir alguna notación. Utilizaremos la notación z_α para designar el número tal que $\Pr\{Z < z_\alpha\} = 1 - \alpha$ y $\Pr\{Z > z_\alpha\} = \alpha$, como muestra la Figura 4.3.12. Así pues, $z_{0,025} = 1,96$.

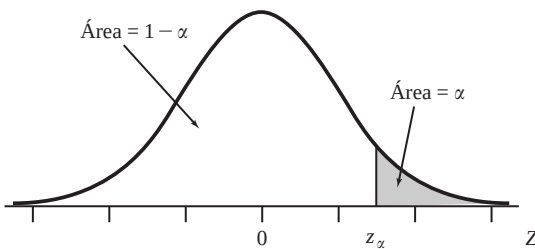


Figura 4.3.11 Área bajo una curva normal por encima de 1,96

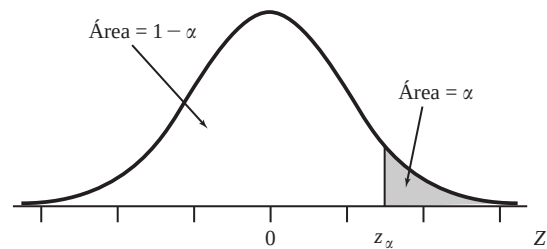


Figura 4.3.12 Área bajo una curva normal por encima de α

A menudo es necesario determinar un valor z_α cuando deseamos determinar un *percentil* de una distribución normal. Los percentiles de una distribución dividen dicha distribución en 100 partes iguales, lo mismo que los cuartiles la dividen en cuatro partes iguales [de las raíces latinas *centum* («100») y *quartus* («cuarto»)]. Por ejemplo, supongamos que deseamos calcular el percentil 70 de una distribución normal tipificada. Esto significa que deseamos obtener el número $z_{0,30}$ que divide la distribución normal tipificada en dos partes: hacia abajo 70 % y hacia arriba el 30 %. Como ilustra la Figura 4.3.13, necesitamos buscar en la Tabla 3 un área de 0,7000. El valor más cercano es un área de 0,6985, correspondiente a un valor de z de 0,52. Por tanto, $z_{0,30} = 0,52$.

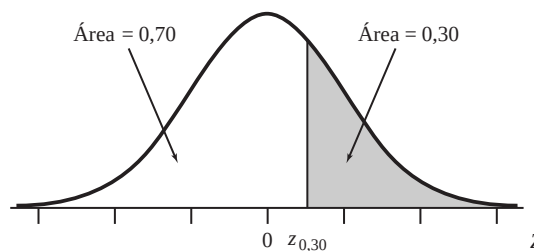


Figura 4.3.13 Determinación del percentil 70 de una distribución normal

Ejemplo 4.3.2 Longitudes de peces

- (a) Supongamos que deseamos obtener el percentil 70 de la distribución de longitudes de peces del Ejemplo 4.3.1. Denominamos y^* a dicho percentil 70. Por definición, y^* es el valor tal que el 70 % de las longitudes de los peces son menores que y^* y el 30 % son mayores, como se ilustra en la Figura 4.3.14.

Para obtener y^* , utilizamos el valor de $z_{0,30} = 0,52$ que acabamos de determinar. Seguidamente, convertimos este valor a la escala Y. Sabemos que si nos dieran el valor de y^* , podríamos transformarlo en una normal tipificada (escala z) y el resultado sería 0,52. Por tanto, utilizando la fórmula de tipificación obtenemos la ecuación

$$0,52 = \frac{y^* - 54}{45}$$

que se puede resolver obteniéndose $y^* = 54 + 0,52 \times 4,5 = 56,3$. El percentil 70 de la distribución de la longitud de peces es 56,3 mm.

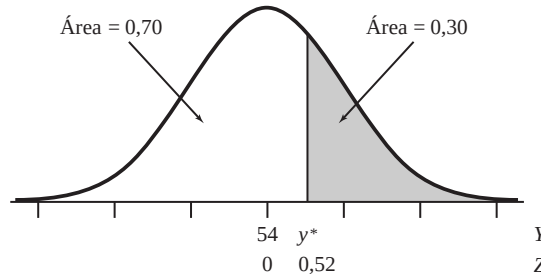


Figura 4.3.14 Determinación del percentil 70 de una distribución normal, Ejemplo 4.3.2(a)

- (b) Supongamos que deseamos calcular el percentil 20 de la distribución de la longitud de peces del Ejemplo 4.3.1. Denominamos y^* a dicho percentil 20. Por definición, y^* es el valor tal que el 20 % de las longitudes de los peces son menores que y^* y el 80 % son mayores, como se ilustra en la Figura 4.3.15.

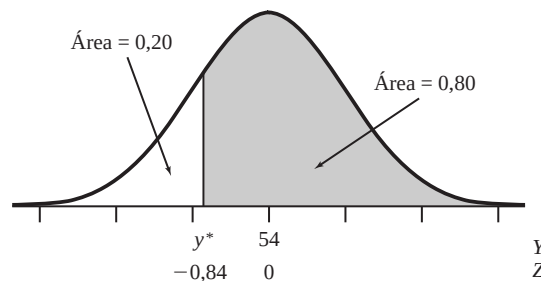


Figura 4.3.15 Determinación del percentil 20 de una distribución normal, Ejemplo 4.3.2(b)

Para obtener y^* , determinamos el valor de $z_{0,80}$, que es el percentil 20 en la escala Z. Como ilustra la Figura 4.3.15, necesitamos buscar en la Tabla 3 un área de 0,2000. El valor más cercano es un área de 0,2005, correspondiente a $z = -0,84$. El paso siguiente es convertir este valor de z a la escala Y. Utilizando la fórmula de tipificación obtenemos la ecuación

$$-0,84 = \frac{y^* - 54}{45}$$

que se puede resolver obteniéndose $y^* = 54 - 0,84 \times 4,5 = 50,2$. El percentil 20 de la distribución de la longitud de peces es 50,2 mm. ■

Consejo para la solución de problemas Al resolver problemas que requieran el uso de la Tabla 3, un dibujo de la distribución (como en las Figuras 4.3.7-4.3.10 y 4.3.14-4.3.15) es una ayuda muy práctica para pensar correctamente.

Aunque la Tabla 3 es práctica para realizar el tipo de cálculos presentados previamente, se puede utilizar también *software* estadístico para calcular directamente probabilidades normales sin la necesidad de ninguna tipificación.

Ejercicios 4.3.1-4.3.16

4.3.1 Suponga que una cierta población de observaciones está distribuida normalmente. Indique qué porcentaje de las observaciones de la población

- (a) Está dentro de $\pm 1,5$ desviaciones típicas alrededor de la media.
- (b) Está más de 2,5 desviaciones típicas por encima de la media.
- (c) Está más de 3,5 desviaciones típicas separado de la media (por encima o por debajo).

4.3.2

- (a) ¿Cuántas desviaciones típicas por encima de la media está el percentil 90 de una distribución normal?
- (b) ¿Cuántas desviaciones típicas por debajo de la media está el percentil 10 de una distribución normal?

4.3.3 Los pesos del cerebro de una cierta población de varones suecos adultos siguen aproximadamente una distribución normal de media 1.400 g y desviación típica 100 g⁶. Indique qué porcentaje de los pesos del cerebro

- (a) Es menor o igual que en 1.500 g.
- (b) Está entre 1.325 y 1.500 g.
- (c) Es mayor o igual que 1.325 g.
- (d) Es mayor o igual que 1.475 g.
- (e) Está entre 1.475 y 1.600 g.
- (f) Está entre 1.200 y 1.325 g.

4.3.4 Sea Y la variable que representa un peso del cerebro aleatoriamente escogido de la población del Ejercicio 4.3.3. Calcule

- (a) $\Pr\{Y \leq 1.325\}$.
- (b) $\Pr\{1.475 \leq Y \leq 1.600\}$.

4.3.5 En un experimento de agricultura se plantó un gran campo uniforme con una única variedad de trigo. El campo se dividió en muchas parcelas (cada una de ellas de 7×100 pies) y se midió la producción de grano (en libras) en cada parcela. La producción de las parcelas siguió aproximadamente una distribución normal de media 88 lb y desviación típica 7 lb⁷. Indique qué porcentaje de las cosechas de las parcelas

- (a) Fue mayor o igual que 80 lb.
- (b) Fue mayor o igual que 90 lb.
- (c) Fue menor o igual que 75 lb.
- (d) Estuvo entre 75 y 90 lb.
- (e) Estuvo entre 90 y 100 lb.
- (f) Estuvo entre 75 y 80 lb.

4.3.6 En relación con el Ejercicio 4.3.5, sea Y la variable que representa la producción de una parcela del campo escogida aleatoriamente. Calcule

- (a) $\Pr\{Y > 90\}$.
- (b) $\Pr\{75 < Y < 90\}$.

4.3.7 Considere una distribución normal tipificada, Z . Calcule

- (a) $z_{0,10}$
- (b) $z_{0,25}$
- (c) $z_{0,05}$
- (d) $z_{0,01}$

4.3.8 En la distribución de la producción de trigo del Ejercicio 4.3.5, calcule

- (a) El percentil 65.
- (b) El percentil 35.

4.3.9 Los niveles de colesterol en suero en niños entre 12 y 14 años siguen una distribución normal de media 162 mg/dl y desviación típica 28 mg/dl. Indique qué porcentaje de niños entre 12 y 14 años de edad tienen valores de colesterol en suero

- (a) Mayores o iguales que 171.
- (b) Menores o iguales que 143.
- (c) Menores o iguales que 194.
- (d) Mayores o iguales que 105.
- (e) Entre 166 y 194.
- (f) Entre 105 y 138.
- (g) Entre 138 y 166.

4.3.10 En relación con el Ejercicio 4.3.9, suponga que se escoge aleatoriamente un niño de 13 años y sea Y su valor de colesterol en suero. Calcule

- (a) $\Pr\{Y \geq 166\}$.
- (b) $\Pr\{166 < Y < 194\}$.

4.3.11 Para la distribución de colesterol en suero del Ejercicio 4.3.9, calcule

- (a) El percentil 80.
- (b) El percentil 20.

4.2.12 Cuando se cuentan glóbulos rojos utilizando un cierto contador electrónico, la desviación típica de medidas repetidas del mismo espécimen de sangre es aproximadamente de 0,8 % del valor verdadero, y la distribución de las medidas repetidas es aproximadamente normal⁸. Por ejemplo, esto significa que si el valor verdadero es de 5.000.000 células/mm³, entonces la DT es de 40.000.

- (a) Si el verdadero valor de la medida de glóbulos rojos de un cierto espécimen es de 5.000.000 células/mm³, ¿cuál es la probabilidad de que la medida dé un valor entre 4.900.000 y 5.100.000?

- (b) Si el verdadero valor de la medida de glóbulos rojos con cierto espécimen es μ , ¿cuál es la probabilidad de que la medida de un valor entre $0,98\mu$ y $1,02\mu$?
- (c) El laboratorio de un hospital realiza medidas en muchos especímenes cada día. ¿Para qué porcentaje de esos especímenes la medida de glóbulos rojos difiere del valor correcto en un 2 % o más?

4.3.13 Se ha encontrado que la cantidad de crecimiento, durante un periodo de 15 días, de una población de plantas de girasol sigue una distribución normal de media 3,18 cm y desviación típica 0,53 cm⁹. Indique qué porcentaje de plantas crecen

- (a) 4 cm o más.
 (b) 3 cm o menos.
 (c) Entre 2,5 y 3,5 cm.

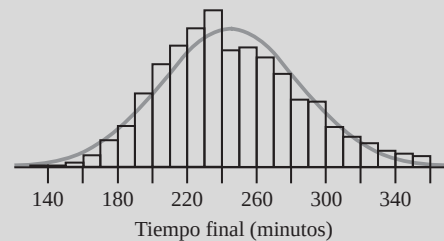
4.3.14 En relación con el Ejercicio 4.3.13, indique en qué intervalo está en 90 % central de todos los valores de crecimiento.

4.3.15 En la distribución del crecimiento de plantas de girasol del Ejercicio 4.3.13, ¿cuál es el percentil 25?

4.3.16 Muchas ciudades patrocinan maratones cada año. El histograma siguiente muestra la distribución de tiempos que

tardaron 10.002 corredores en completar la maratón de Roma en 2008, con una curva normal superpuesta. El corredor más rápido completó la carrera de 26,3 millas en 2 horas y 9 minutos, o 129 minutos. El tiempo medio fue de 245 minutos y la desviación típica de 40 minutos. Utilice la curva normal para responder a las siguientes preguntas¹⁰.

- (a) ¿Qué porcentaje de tiempos fueron superiores a 200 minutos?
- (b) ¿Cuál es el percentil 60 de los tiempos?
- (c) Advierta que la aproximación mediante la curva normal es bastante buena excepto alrededor de la marca de 240 minutos. ¿Cómo puede explicar el comportamiento anómalo de la distribución?



4.4 Evaluación de la normalidad

Muchos procedimientos estadísticos se basan en tener datos de una población normal. En esta sección consideraremos formas de evaluar si es razonable utilizar un modelo de curva normal para un conjunto de datos, y si no, cómo podríamos proceder.

Recordemos de la Sección 4.3 que si la variable Y sigue una distribución normal, entonces:

- aproximadamente el 68 % de los valores de y están dentro de ± 1 DT alrededor de la media.
 aproximadamente el 95 % de los valores de y están dentro de ± 2 DT alrededor de la media.
 aproximadamente el 99,7 % de los valores de y están dentro de ± 3 DT alrededor de la media.

Podemos utilizar entonces estos hechos para comprobar el ajuste de un modelo de curva normal a un conjunto de datos.

Ejemplo 4.4.1 Colesterol en suero

Para los datos de colesterol en suero del Ejemplo 4.1.1, la media muestral es 162 y la DT muestral es 28. El intervalo «media $\pm$ DT» es

$$(162 - 28, 162 + 28) \text{ o } (134, 190)$$

Este intervalo contiene 509 de las 727 observaciones, o el 70,0 % de los actos. Asimismo, el intervalo

$$(162 - 2 \times 28, 162 + 2 \times 28) \text{ es } (106, 218)$$

que contiene 685, o el 94,2 %, de las 727 observaciones. Finalmente, el intervalo

$$(162 - 3 \times 28, 162 + 3 \times 28) \text{ es } (78, 246)$$

que contiene 724, con 99,6 %, de las 727 observaciones. Los tres porcentajes observados

$$70,0 \%, 94,2 \% \text{ y } 99,6 \%$$

se ajustan bastante bien a los porcentajes teóricos de

68 %, 95 % y 99,7%

Este ajuste avala la afirmación de que los niveles de colesterol en suero para niños entre 12 y 14 años tienen una distribución normal, y refuerza la evidencia visual de la Figura 4.1.1. ■

Ejemplo 4.4.2 Contenido de humedad

Se midió el contenido de humedad en 83 frutas de agua dulce¹¹. La Figura 4.4.1 muestra que esta distribución está fuertemente sesgada hacia la izquierda. La media muestral en estos datos es 80,7 y la DT muestral es 12,7. El intervalo

$$(80,7 - 12,7, 80,7 + 12,7)$$

contiene 70, o el 84,3 %, de las 83 observaciones. El intervalo

$$(80,7 - 2 \times 12,7, 80,7 + 2 \times 12,7)$$

contiene 78, o el 94,0 %, de las 83 observaciones. Finalmente, el intervalo

$$(80,7 - 3 \times 12,7, 80,7 + 3 \times 12,7)$$

contiene 80, o el 96,4 % de las 83 observaciones. Los tres porcentajes

84,3 %, 94,0 % y 96,4 %

difieren de los porcentajes teóricos de

68 %, 95 % y 99,7 %

debido a que la distribución está lejos de tener forma de campana. Esto refuerza la evidencia visual de la Figura 4.4.1. ■

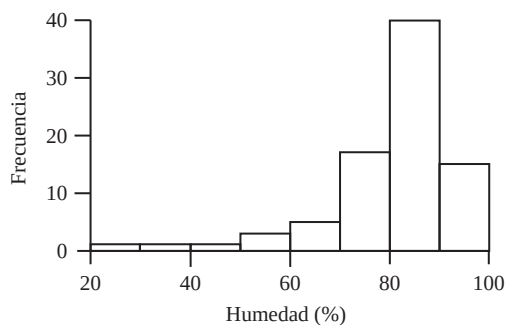


Figura 4.4.1 Contenido de humedad de fruta de agua dulce

Gráficas de probabilidad normal

Una **gráfica de probabilidad normal** es una gráfica estadística especial que se utiliza para evaluar la normalidad. Presentamos esta herramienta estadística mediante un ejemplo utilizando las alturas (en pulgadas) de una muestra de 11 mujeres, ordenadas de menor a mayor:

61, 62,5, 63, 64, 64,5 65, 66,5, 67, 68, 68,5, 70,5

Basándose en estos datos, ¿tiene sentido utilizar una curva normal para modelar la distribución de las alturas de las mujeres? La Figura 4.4.2 es un histograma de los datos con una curva normal superpuesta, utilizando la media muestral de 65,5 y la desviación típica muestral de 2,9 como parámetros de la curva normal. Este histograma es bastante simétrico, pero cuando tenemos una muestra pequeña, puede ser difícil decir la forma de la distribución poblacional observando el histograma.

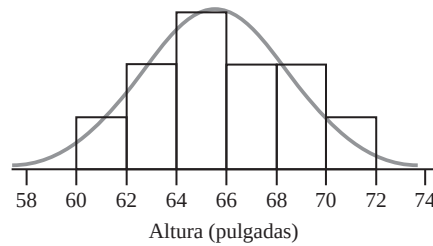


Figura 4.4.2 Histograma de las alturas de 11 mujeres

Como a menudo es difícil examinar visualmente un histograma y decidir si tiene forma de campana o no, se ha desarrollado una gráfica visualmente más simple, la gráfica de la probabilidad normal*. Una gráfica de probabilidad normal es un diagrama de dispersión que compara nuestros valores de los datos observados con los valores que habría que esperar ver si la población fuera normal. Si los datos provienen de una población normal, los puntos de esta gráfica deberían seguir una línea recta, que es mucho más fácil de reconocer visualmente que una forma de campana de un histograma irregular. Como muchos procedimientos estadísticos se basan en la condición de que los datos provienen de una población normal, es importante ser capaces de evaluar la normalidad.

Cómo funcionan las gráficas de probabilidad normal

En los Ejemplos 4.4.1 y 4.4.2 comparamos la proporción observada de los datos que estaban dentro de 1, 2 y 3 DT alrededor de la media y después comparamos esos valores con las proporciones que cabría esperar encontrar si los datos provinieran de una población normal. Es natural considerar estos intervalos, pero podríamos considerar otros intervalos también. Por ejemplo, se puede esperar que aproximadamente el 86,6 % de datos normales estén dentro de 1,5 DT alrededor de la media y que el 96,4 % esté dentro de 2,1 DT alrededor de la media†. Incluso podríamos considerar intervalos unilaterales. Por ejemplo, se puede esperar que el 84,1 % de valores de datos normales sean menores que la media más 1 DT.

En vez de concentrarnos en comparar porcentajes, podemos centrarnos en comparar los valores reales observados de las alturas de mujeres que esperaríamos ver si los datos provinieran de una población normal. Por ejemplo, la mujer más baja de nuestra muestra tiene una altura de 61 pulgadas. Es decir, 1/11 (o 0,0909) de la muestra tiene valores de 61 pulgadas o menos. Si las alturas de las mujeres siguieran realmente una distribución normal, de media 65,5 y la desviación típica 2,9, entonces se puede esperar que el percentil 9,09 sea $\mu + z_{(1-0,0909)}\sigma = 65,5 - 1,34 \times 2,9$, o 61,6 pulgadas. Este valor es cercano al valor observado de 61 pulgadas. Podríamos repetir este tipo de cálculo para cada uno de los 11 valores de los datos observados. Una gráfica de probabilidad normal proporciona una comparación visual de estos valores.

El primer paso para crear una gráfica de probabilidad normal, por tanto, es calcular los percentiles muestrales. El Ejemplo 4.4.3 presenta la forma de realizar este cálculo, que en general es realizado por los paquetes de *software* estadístico.

Ejemplo 4.4.3 Altura de 11 mujeres

Al ordenar los datos de menor a mayor observamos que 1/11 (=9,1 %) de nuestra muestra mide 61 pulgadas o menos, 2/11 (=18,2 %) mide 62,5 pulgadas o menos, ... 10/11 (90,9 %) mide 68,5 pulgadas o menos y 11/11 (100 %) mide 70,5 pulgadas o menos. Desafortunadamente, el cálculo de porcentajes de esta forma simplista (es decir, $100 \times i/n$, siendo i el número de la observación ordenada) produce algunos estimadores poblacionales inverosímiles. Por ejemplo, no parece razonable pensar que el 100 % de la *población* mide 70,5 pulgadas o menos cuando, después de todo, solo estamos observando una pequeña muestra. Una muestra mayor posiblemente permitiría observar algunas mujeres más altas. Para corregir esto, se puede calcular un porcentaje alternativo y más razonable para

* Aunque visualmente simple, la función de estas gráficas es compleja y en general se realiza utilizando *software* estadístico.

† Estos valores se pueden verificar utilizando las técnicas de la Sección 4.3.

cada valor de datos como $100(i - \frac{1}{2})/n$, siendo i el índice del valor de datos en la lista ordenada*. Estos percentiles ajustados se muestran en la Tabla 4.4.1. Nótese que estos valores realmente no dependen de los datos observados, ya que solo dependen del número de valores de datos en la muestra. ■

Tabla 4.4.1 Cálculo de índices y percentiles de las alturas de 11 mujeres

i	1	2	3	4	5	6	7	8	9	10	11
Altura observada	61,0	62,5	63,0	64,0	64,5	65,0	66,5	67,0	68,0	68,5	70,5
Percentil $100(i/11)$	9,09	18,18	27,27	36,36	45,45	54,55	63,64	72,73	81,82	90,91	100,00
Pertencil ajustado $100(i - \frac{1}{2})/11$	4,55	13,64	22,73	31,82	40,91	50,00	59,09	68,18	77,27	86,36	95,45

Una vez que tenemos los percentiles ajustados, calculamos las correspondientes puntuaciones z utilizando la Tabla 3 o un computador. Después, con estas puntuaciones z calculamos las alturas teóricas: $\mu + z \times \sigma$ como en el Ejemplo 4.4.4.

Ejemplo 4.4.4 Altura de 11 mujeres

El percentil ajustado de la mujer más baja es 4,55 %. El valor de la puntuación z correspondiente es $z_{(1-0,0455)} = z_{0,9545} = -1,69$. En este ejemplo, la media muestral y la desviación típica muestral son 65,5 y 2,9, respectivamente, por lo que la altura esperada de la mujer más baja en la muestra de las 11 mujeres de una población normal tiene un valor de $65,5 - 1,69 \times 2,9 = 60,6$ pulgadas. Las puntuaciones z y las alturas teóricas para esta mujer y las restantes 10 mujeres aparecen en la Tabla 4.4.2.

Tabla 4.4.2 Cálculo de puntuaciones z y alturas teóricas de 11 mujeres

i	1	2	3	4	5	6	7	8	9	10	11
Altura observada	61,0	62,5	63,0	64,0	64,5	65,0	66,5	67,0	68,0	68,5	70,5
Percentil ajustado $100(i - \frac{1}{2})/11$	4,55	13,64	22,73	31,82	40,91	50,00	59,09	68,18	77,27	86,36	95,45
z	-1,69	-1,10	-0,75	-0,47	-0,23	0,00	0,23	0,47	0,75	1,10	1,69
Altura teórica	60,6	62,3	63,4	64,1	64,8	65,5	66,2	66,9	67,6	68,7	70,4

Seguidamente, dibujando las alturas observadas en función de las alturas teóricas en un diagrama de dispersión, como se muestra en la Figura 4.4.3, podemos comparar visualmente los valores. En este caso, nuestra gráfica tiene un aspecto bastante lineal, lo que sugiere que en general los valores observados están de acuerdo con los valores teóricos (es decir, que el modelo normal proporciona una aproximación razonable de los datos). Si los datos no se ajustan con un modelo normal, entonces la gráfica mostrará fuertes patrones no lineales, como curvaturas o formas de S.

Debido a la correspondencia uno a uno entre las puntuaciones z y los valores teóricos, no es común poner ambos conjuntos de valores en el eje x como en la Figura 4.4.3. Tradicionalmente, solo se muestran las puntuaciones $z^{\dagger}$. ■

Toma de decisiones sobre normalidad

Por supuesto, incluso cuando tomamos una muestra de una población perfectamente normal, debemos esperar que haya alguna variabilidad entre la muestra que obtenemos y las puntuaciones normales teóricas. La Figura 4.4.4 muestra

* Diferentes paquetes *software* pueden calcular estas proporciones de manera diferente y pueden también modificar la fórmula basada en el tamaño de la muestra. La fórmula anterior es la utilizada por el paquete de *software R* cuando $n > 10$.

[†] Algunos programas de *software* crean gráficas de probabilidad normal con las puntuaciones normales en el eje vertical y los valores observados en el eje horizontal.

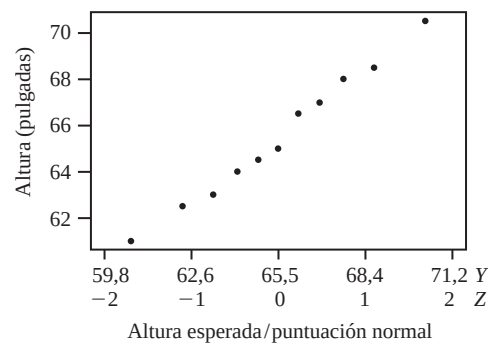


Figura 4.4.3 Gráfica de probabilidad normal de las alturas de 11 mujeres

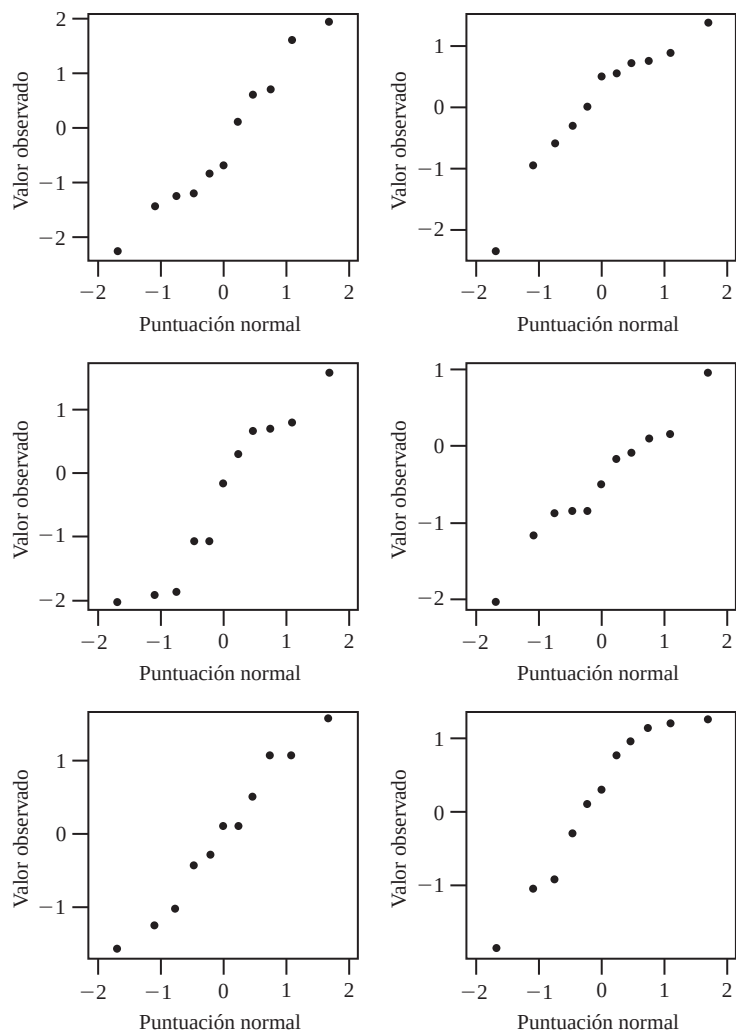


Figura 4.4.4 Gráficas de probabilidad normal para datos normales

seis gráficas de probabilidad normal basadas en muestras tomadas de una distribución $N(0, 1)$. Obsérvese que las seis gráficas muestran un patrón general lineal. Es cierto que existe una cierta «curvatura» en alguna de las gráficas, pero lo importante en cada una de esas gráficas es que podemos dibujar una recta que captura la tendencia de la mayoría de los puntos, con poca desviación con respecto a dicha recta, incluso en los extremos.

Si los puntos de una gráfica de probabilidad normal no están más o menos a lo largo de una línea recta, esto es una indicación de que los datos no provienen de una población normal. Por ejemplo, si la parte alta de la gráfica se curva, esto significa que los valores de y en el extremo superior de la distribución son demasiado grandes para que la distribución tenga forma de campana. Es decir, la distribución está sesgada hacia la derecha o tiene *outliers* grandes, como se muestra en la Figura 4.4.5.

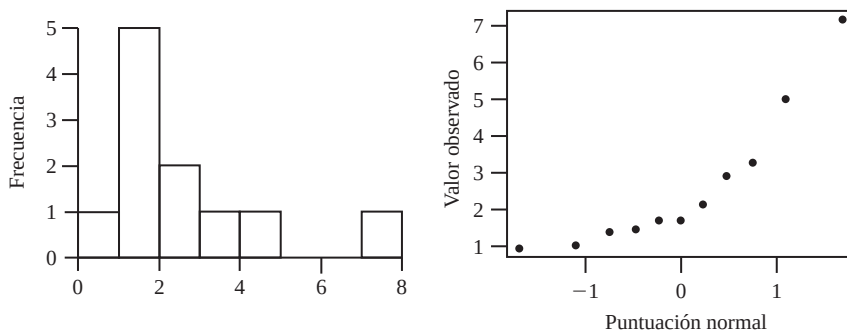


Figura 4.4.5 Histograma y gráfica de probabilidad normal de una distribución que está sesgada hacia la derecha

Si la parte baja de la gráfica se curva hacia abajo, esto significa que los valores de y en el extremo inferior de la distribución son demasiado pequeños para que la distribución tenga forma de campana. Es decir, la distribución está sesgada hacia la izquierda o tiene *outliers* pequeños. La Figura 4.4.6 muestra la distribución del contenido de humedad en la fruta de agua dulce del Ejemplo 4.4.2, que está fuertemente sesgada hacia la izquierda.

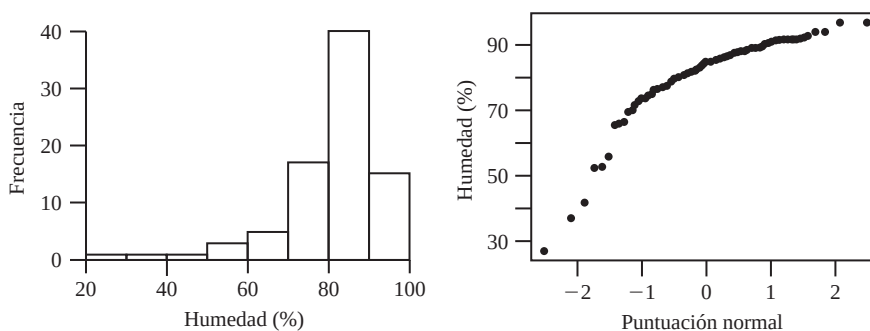


Figura 4.4.6 Histograma y gráfica de probabilidad normal de una distribución que está sesgada hacia la izquierda

Si la distribución tiene una cola muy larga hacia la izquierda y una cola larga hacia la derecha, al compararla con la curva normal la gráfica de probabilidad normal tendrá una forma parecida a una S. La Figura 4.4.7 muestra una distribución de ese tipo.

Algunas veces un mismo valor aparece repetidamente en una muestra, debido al redondeo del proceso de medida. Esto produce una *granularidad* en la gráfica de probabilidad normal, como la Figura 4.4.8, pero esto no evita que sigamos deduciendo que la distribución subyacente es normal.

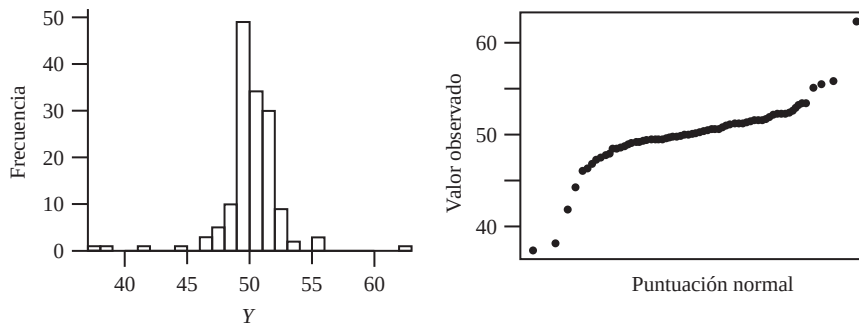


Figura 4.4.7 Histograma y gráfica de probabilidad normal de una distribución que tiene colas largas

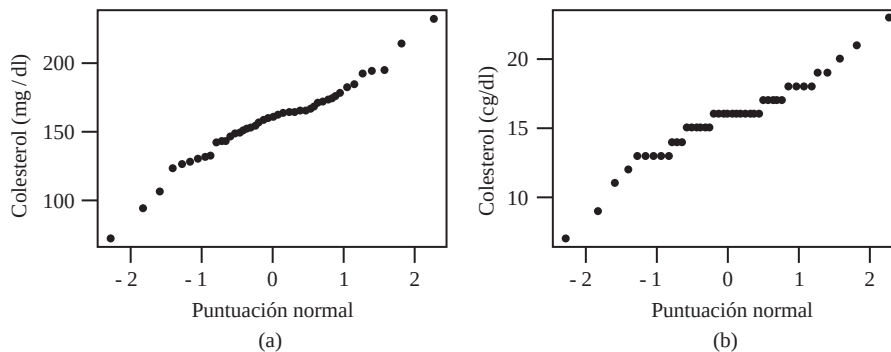


Figura 4.4.8 Gráficas de probabilidad normal de valores de colesterol de 50 niños entre 12 y 14 años medidos con una precisión de (a) mg/dl y (b) cg/dl

Transformaciones para datos no normales

Una gráfica de probabilidad normal puede ayudarnos a verificar si los datos provienen o no de una distribución normal. Algunas veces, un histograma de una gráfica de probabilidad normal nos muestran que nuestros datos no son normales, pero una transformación de dichos datos puede producir una curva simétrica con forma de campana. En esta situación, puede ser deseable transformar los datos y continuar nuestro análisis en la nueva (transformada) escala.

Ejemplo 4.4.5 Crecimiento de lentejas

El histograma y la gráfica de probabilidad normal de la Figura 4.4.9 muestran la distribución de la velocidad de crecimiento, en cm por día, de una muestra de 47 plantas de lenteja¹². Esta distribución está sesgada hacia la derecha. Si tomamos el logaritmo de cada observación, obtenemos una distribución mucho más simétrica. Las gráficas de la Figura 4.4.10 muestran que, en escala logarítmica, la distribución de la velocidad de crecimiento es aproximadamente normal. (En la Figura 4.4.10 se utiliza el logaritmo en base 10, $\log_{10}$, pero podríamos utilizar cualquier base, como logaritmo natural, $\log_e = \ln$, y el efecto sobre la forma de la distribución sería el mismo). ■

En general, si la distribución está sesgada hacia la derecha, entonces debería considerarse alguna de las siguientes transformaciones: $\sqrt{Y}$, $\log Y$, $1/\sqrt{Y}$, $1/Y$. Estas transformaciones encogerán la cola larga hacia la derecha y estirarán la cola corta hacia la izquierda, haciendo que la distribución se haga más simétrica. Cada una de ellas es más drástica que la que le precede. Así, una transformación de raíz cuadrada cambiará una distribución ligeramente sesgada en una distribución simétrica, pero puede ser necesaria una transformación logarítmica si la distribución está más fuertemente sesgada, y así sucesivamente. Por ejemplo, vimos en el Ejemplo 2.7.6 cómo una transformación de tipo raíz cuadrada encoge una cola larga hacia la derecha y como una transformación logarítmica encoge la cola hacia la derecha todavía más. Si la distribución de una variable Y está sesgada hacia la izquierda, entonces elevar Y a una potencia mayor que 1 puede ser útil.

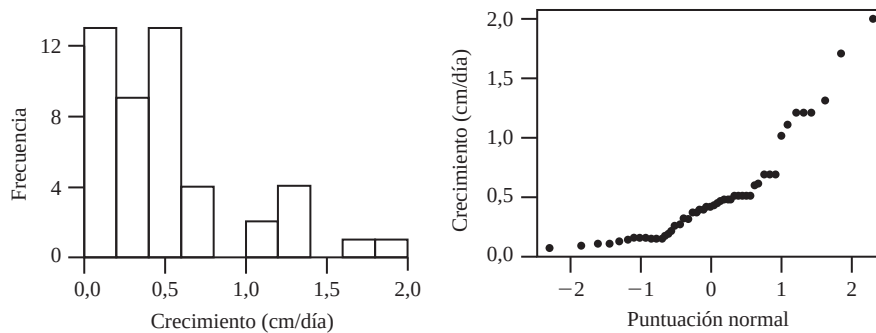


Figura 4.4.9 Histograma y gráfica de probabilidad normal de las velocidades de crecimiento de 47 plantas de lenteja

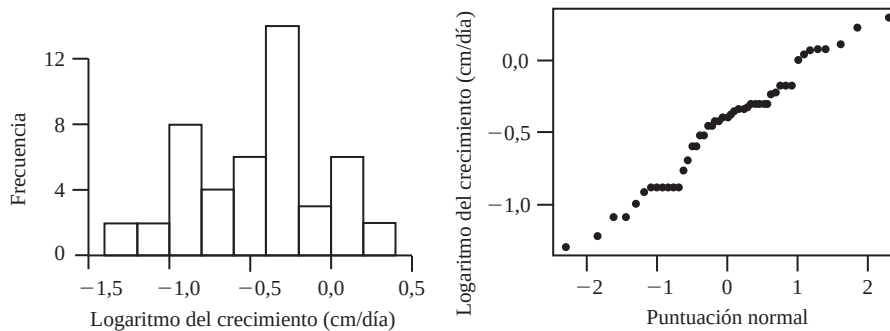


Figura 4.4.10 Histograma y gráfica de probabilidad normal del logaritmo de las velocidades de crecimiento de 47 plantas de lenteja

Una medida objetiva de no normalidad: la prueba de Shapiro-Wilk (opcional)

Aunque las gráficas de probabilidad normal son mejores que los histogramas para evaluar visualmente desviaciones de la normalidad, nuestra percepción visual es todavía subjetiva. Los datos que aparecen en las curvas de probabilidad de la Figura 4.4.4 provienen de una población normal, pero para ojos no entrenados (e incluso para algunos entrenados) algunas de las gráficas se podrían interpretar como no normales. La **prueba de Shapiro-Wilk** es un procedimiento estadístico que evalúa numéricamente la evidencia de ciertos tipos de no normalidad en los datos. Como en el caso de las gráficas de probabilidad normal, la mecánica del procedimiento es compleja, pero afortunadamente muchos paquetes de *software* estadístico realizan esta o pruebas similares de normalidad*.

La salida de una prueba de Shapiro-Wilk es un P -valor[†] y se interpreta como sigue:

P -valor $< 0,001$	Evidencia muy fuerte de no normalidad
P -valor $< 0,01$	Evidencia fuerte de no normalidad
P -valor $< 0,05$	Evidencia moderada de no normalidad
P -valor $< 0,10$	Evidencia débil de no normalidad
P -valor $\geq 0,10$	No existe evidencia convincente de no normalidad

* Las pruebas de Ryan-Joiner, Anderson-Darling y Kolmogorof-Smirnoff son otras pruebas de no normalidad que se encuentran comúnmente en paquetes de *software* estadístico.

[†] Como veremos con mucho más detalle en el Capítulo 7, un P -valor no es único para probar la normalidad. En una prueba de todos los tipos de hipótesis, el peso de la evidencia para la hipótesis en cuestión (en este caso, la prueba de Shapiro-Wilk, la hipótesis es que los datos son no normales) se puede indicar utilizando este término. P -valores pequeños se interpretan como una evidencia para la hipótesis en cuestión.

El Ejemplo 4.4.6 ilustra la prueba de Shapiro-Wilk para los datos de crecimiento de lentejas del Ejemplo 4.4.5.

Ejemplo 4.4.6 Crecimiento de lentejas

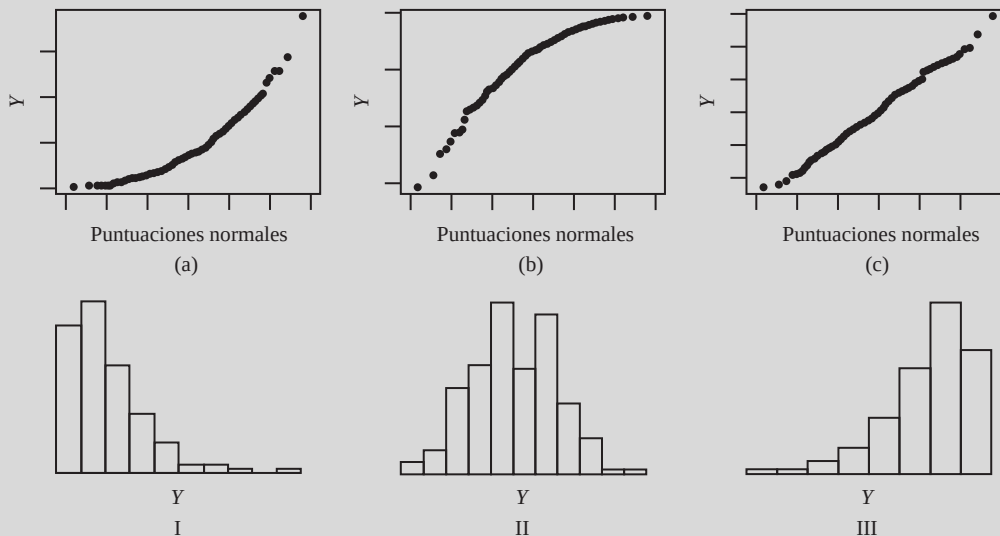
Para los datos de crecimiento de lentejas no transformados de la Figura 4.4.9, el P -valor (que proporciona el paquete de *software* estadístico *R*) para la prueba de Shapiro-Wilk es 0,000006. Por tanto, hay una evidencia muy fuerte de que el crecimiento de lentejas no sigue una distribución normal. Sin embargo, para los datos transformados de la Figura 4.4.10, el P -valor para la prueba de Shapiro-Wilk es 0,2090, lo que indica que no existe evidencia convincente de la no normalidad de los datos de crecimiento transformados logarítmicamente. ■

Advertencia El uso de este procedimiento de prueba y del P -valor es algo así como el uso de una «luz de aviso del motor» de un coche. Cuando el P -valor es pequeño, hay una indicación de no normalidad. Esto es como cuando la luz de aviso del motor se enciende. Nos hacemos a un lado y evaluamos la situación. Asimismo, como veremos en futuros capítulos, cuando tenemos datos no normales, tenemos que verificar cuidadosamente cómo procedemos con nuestro análisis. Por otra parte, cuando el P -valor no es pequeño ($\geq 0,10$) no tenemos evidencia de no normalidad. Esto es similar a que nuestra luz de aviso del motor está apagada: continuamos conduciendo sin preocuparnos, *pero* esto no garantiza que nuestro coche esté perfectamente. Nuestro coche podría averiarse en cualquier momento. Por supuesto, si estuviéramos constantemente preocupados con nuestro coche incluso cuando la luz de aviso del motor está apagada, nos encontraríamos perpetuamente paralizados y apartados en el arcén. Análogamente, cuando el P -valor de las pruebas de Shapiro-Wilk no es pequeño (la luz está apagada), esto solo significa que no existe evidencia convincente de la no normalidad. Esto no garantiza que la población sea, de hecho, normal.

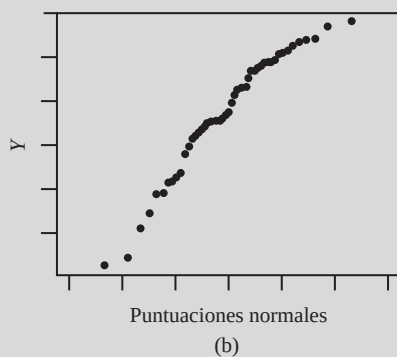
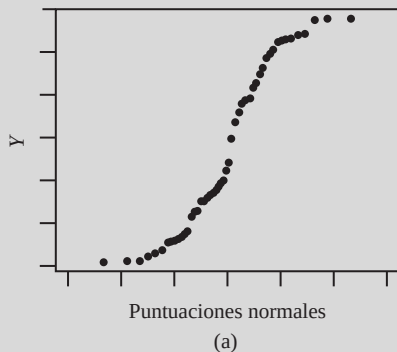
Ejercicios 4.4.1-4.4.8

4.4.1 En el Ejemplo 4.1.2 se indicó que el espesor de la cáscara de una población de huevos seguía una distribución normal de media $\mu = 0,38$ mm y desviación típica $\sigma = 0,03$ mm. Utilice la regla del 68 % – 95 % – 99,7 % para determinar intervalos, centrados en la media, que incluyan el 68 %, el 95 % y el 99,7 % del espesor de las cáscaras en la distribución.

4.4.2 Las siguientes tres gráficas de probabilidad normal, (a), (b) y (c), se han generado a partir de las distribuciones que se muestran en los histogramas I, II y III. ¿Qué gráficas de probabilidad normal corresponden a cada histograma? ¿Cómo lo sabe?



4.4.3 Para cada uno de las siguientes gráficas de probabilidad normal, dibuje los correspondientes histogramas de los datos.

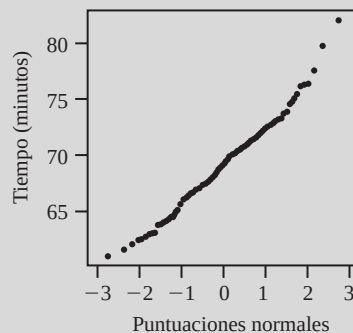


4.4.4 La pluviosidad media diaria entre el 1 de enero de 2007 y el 1 de enero de 2009 en Pismo Beach, California, fue de 0,02 pulgadas con una desviación típica de 0,11 pulgadas. Basándose en esta información, ¿piensa que es razonable suponer que la pluviosidad diaria en Pismo Beach sigue una distribución normal? Explique su respuesta. (*Sugerencia:* piense en los posibles valores de la pluviosidad diaria)¹³.

4.4.5 La media de las temperaturas máximas diarias del día 1 de febrero en Juneau, Alaska, entre 1945 y 2005 fue de 1,1 °C y la desviación típica de 1,9 °C¹⁴.

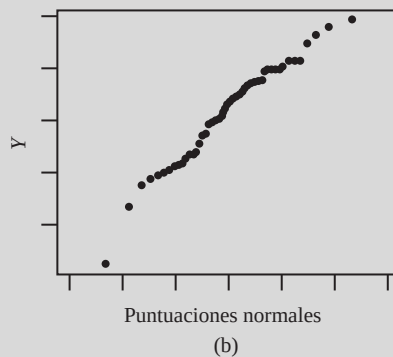
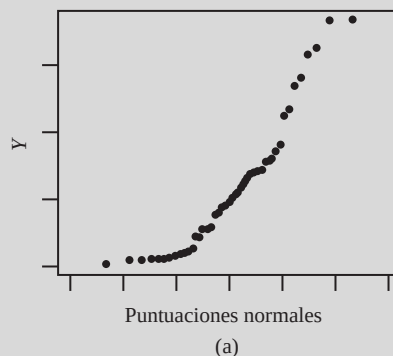
- (a) Basándose en esta información, ¿piensa que es razonable suponer que las temperaturas máximas diarias el día 1 de febrero en Juneau, Alaska, siguen una distribución normal? Explique su respuesta.
- (b) ¿Proporciona esta información evidencia convincente de que las temperaturas máximas diarias el día 1 de febrero en Juneau, Alaska, siguen una distribución normal? Explique su respuesta.

4.4.6 Las siguientes gráficas de probabilidad normal se han creado a partir de los tiempos empleados por 166 ciclistas en completar la etapa 11 contra reloj desde Grenoble hasta Chamrousse, Francia, en la carrera ciclista Tour de Francia de 2001.



- (a) Considere a los corredores más rápidos. ¿Son sus tiempos mejores que, peores que, o aproximadamente iguales a los tiempos que uno esperaría encontrar para los corredores más rápidos si los datos provinieran de una distribución verdaderamente normal?
- (b) Considere a los corredores más lentos. ¿Son sus tiempos mejores que, peores que, o aproximadamente iguales a los tiempos que uno esperaría encontrar para los corredores más lentos si los datos provinieran de una distribución verdaderamente normal?

4.4.7 Los P -valores de la prueba de Shapiro-Wilk para los datos que aparecen en las gráficas de probabilidad (a) y (b) son 0,235 y 0,00015. ¿Qué P -valor corresponde a cada gráfica? ¿En qué basa su decisión?



4.4.8

(a) El P -valor de la prueba de normalidad de Shapiro-Wilk del Ejercicio 4.4.3(b) es 0,039. Utilizando este valor para justificar su respuesta, ¿le parece razonable suponer que estos datos provienen de una población normal?

(b) El P -valor de la prueba de normalidad de Shapiro-Wilk del Ejercicio 4.4.3(c) es 0,770. Utilizando este valor para justificar su respuesta, ¿le parece razonable suponer que estos datos provienen de una población normal?

(c) ¿Demuestra el P -valor del apartado (b) que los datos provienen de una población normal?

4.5 Perspectiva

La distribución normal se denomina también distribución Gaussiana tras los trabajos del matemático alemán K. F. Gauss. El término *normal*, con sus connotaciones de «típico» o «usual», puede ser seriamente engañoso. Consideremos, por ejemplo, un contexto médico donde el significado principal de «normal» es «no anormal». Por tanto, de manera confusa, la frase «la población normal de niveles de colesterol en suero» puede referirse a los niveles de colesterol en personas idealmente «saludables» o se puede referir a una distribución Gaussiana como la del Ejemplo 4.1.1. De hecho, para muchas variables la distribución de la población normal (no enferma) es decididamente no normal (es decir, no Gaussiana).

Los ejemplos de este capítulo han ilustrado un uso de la distribución normal: como aproximación a distribuciones biológicas que aparecen de forma natural. Si la distribución natural se aproxima bien mediante una distribución normal, entonces la media y la desviación típica proporcionan una descripción completa de la distribución. La media es el centro de la distribución: aproximadamente el 68 % de los valores están dentro de la desviación típica alrededor de la media, aproximadamente el 95 % están dentro de 2 desviaciones típicas alrededor de la media, y así sucesivamente.

Como se indicó en la Sección 2.6, los valores del 68 % de 95 % se pueden aplicar aproximadamente incluso a distribuciones que están bastante sesgadas. (Pero si la distribución está sesgada, entonces el 68 % no está repartido simétricamente a ambos lados de la media, y de forma similar ocurre para el 95 %). Sin embargo, los valores anteriores no se pueden aplicar a una distribución (incluso aunque sea simétrica) en la que una o ambas colas son largas y finas (véanse las Figuras 2.2.13 y 2.2.16).

En capítulos posteriores veremos que muchos métodos estadísticos clásicos están específicamente diseñados para, y funcionan mejor con, datos que han sido muestreados de poblaciones normales. Veremos posteriormente que en muchas situaciones prácticas estos métodos también funcionan muy bien en el caso de muestras de poblaciones no normales.

La distribución normal tiene una importancia central a pesar del hecho de que muchas, quizá la mayoría, de las distribuciones biológicas que aparecen de forma natural se podrían describir mejor mediante curvas sesgadas que mediante una curva normal. Un uso importante de la distribución normal no es describir distribuciones naturales, sino describir ciertas distribuciones teóricas, denominadas distribuciones en el muestreo, que se utilizan en el análisis estadístico de los datos. En el Capítulo 5 veremos que muchas distribuciones en el muestreo son aproximadamente normales, incluso aunque los datos subyacentes no lo sean. Es esta propiedad la que hace que la distribución normal sea tan importante en el estudio de la estadística.

Ejercicios suplementarios 4.S.1-4.S.21

4.S.1 La actividad de una cierta enzima se mide contando las emisiones de una molécula marcada radiactivamente. En un espécimen de tejido dado, las emisiones en períodos de 10 segundos consecutivos se pueden considerar (aproximadamente) como observaciones independientes repetidas de una distribución normal¹⁵. Supongamos que la media de una emisión de 10 segundos para un cierto espécimen de tejido es 1.200 y que la desviación típica es 35. Sea Y la variable que indica las emisiones en un periodo de tiempo de 10 segundos elegido aleatoriamente. Calcule

(a) $\Pr\{Y \geq 1.250\}$.

(b) $\Pr\{Y \geq 1.175\}$.

(c) $\Pr\{1.150 \leq Y \leq 1.250\}$.

(d) $\Pr\{1.150 \leq Y \leq 1.175\}$.

4.S.2 Los espesores de la cáscara de los huevos producidos por un gran número de gallinas siguen aproximadamente una distribución normal de media igual a 0,38 mm y desviación típica igual a 0,03 mm (como en el Ejemplo 4.1.2). Calcule el percentil 95 de la distribución del espesor.

4.S.3 En relación con la distribución del espesor de las cáscaras de huevo del Ejercicio 4.S.2, suponga que se define un huevo como de cáscara fina si su espesor es de 0,32 mm o menos.

- (a) ¿Qué porcentaje de los huevos son de cáscara fina?
- (b) Suponga que un gran número de huevos se empaquetan aleatoriamente en cajas de 12. ¿Qué porcentaje de las cajas contendrán al menos un huevo de cáscara fina? (*Sugerencia:* calcule primero el porcentaje de cajas que no contendrán ningún huevo de cáscara fina).

4.S.4 Las alturas de una cierta población de plantas de maíz siguen una distribución normal de media 145 cm y de desviación típica 22 cm¹⁶. Indique qué porcentaje de las alturas de las plantas:

- (a) Mide 100 cm o más.
- (b) Mide 120 cm o menos.
- (c) Tiene una altura entre 120 y 150 cm.
- (d) Tiene una altura entre 100 y 120 cm.
- (e) Tiene una altura entre 150 y 180 cm.
- (f) Mide 180 cm o más.
- (g) Mide 150 cm o menos.

4.S.5 Suponga que se escogen aleatoriamente cuatro plantas de la población de plantas de maíz del Ejercicio 4.S.4. Calcule la probabilidad de que ninguna de las cuatro plantas tenga una altura mayor que 150 cm.

4.S.6 En relación con la población de plantas de maíz del Ejercicio 4.S.4, calcule el percentil 90 de la distribución de alturas.

4.S.7 En la población de plantas de trigo descrita en el Ejercicio 4.S.4, calcule los cuartiles y el intervalo intercuartílico.

4.S.8 Suponga una cierta población de observaciones que está normalmente distribuida.

- (a) Calcule el valor de z^* tal que el 95 % de las observaciones de la población está entre $-z^*$ y $+z^*$ en la escala Z.
- (b) Calcule el valor de z^* tal que el 99 % de las observaciones de la población está entre $-z^*$ y $+z^*$ en la escala Z.

4.S.9 En la actividad de las células nerviosas de una cierta mosca concreta, los intervalos de tiempo entre descargas «de pico» siguen aproximadamente una distribución normal de media 5,6 ms y de desviación típica 0,4 ms (como en el Ejemplo 4.1.3). Sea Y la variable que indica un intervalo entre picos seleccionada aleatoriamente. Calcule

- (a) $\Pr\{Y > 15\}$.
- (b) $\Pr\{Y > 16,5\}$.
- (c) $\Pr\{15 < Y < 16,5\}$.
- (d) $\Pr\{15 < Y < 15,5\}$.

4.S.10 Para la distribución de intervalos de tiempo entre picos descrita en el Ejercicio 4.S.9, calcule los cuartiles y el intervalo intercuartílico.

4.S.11 Entre las mujeres americanas entre 20 y 29 años de edad, el 10 % miden menos de 60,8 pulgadas, el 80 % miden entre 60,8 y 67,6 pulgadas y el 10 % miden más de 67,6 pul-

gadas¹⁷. Suponiendo que la distribución de las alturas se puede aproximar adecuadamente mediante una curva normal, calcule la media y la desviación típica de la distribución.

4.S.12 La puntuación del coeficiente intelectual (CI), medido mediante la prueba de Stanford-Binet, está normalmente distribuida en una cierta población de niños. La puntuación media del CI es de 100 puntos, y la desviación típica es de 16 puntos¹⁸. Indique qué porcentaje de los niños de la población tienen puntuaciones de CI:

- (a) De 140 o más.
- (b) De 80 o menos.
- (c) Entre 80 y 120.
- (d) Entre 80 y 140.
- (e) Entre 120 y 140.

4.S.13 En relación con la distribución de CI del Ejercicio 4.S.12, sea Y la variable que indica la puntuación del CI de un niño elegido aleatoriamente de la población. Calcule $\Pr\{80 \leq Y \leq 140\}$.

4.S.14 En relación con la distribución del CI del Ejercicio 4.S.12, suponga que se escogen aleatoriamente cinco niños de la población. Calcule la probabilidad de que exactamente uno de ellos tenga una puntuación del CI de 80 o menos y los otros cuatro tengan puntuaciones mayores que 80. (*Sugerencia:* calcule primero la probabilidad de que un niño escogido aleatoriamente tenga una puntuación del CI de 80 o menos).

4.S.15 Un cierto ensayo sobre alanina aminotransferasa (ALT) es bastante impreciso. Los resultados de ensayos repetidos sobre un mismo espécimen siguen una distribución normal de media igual a la verdadera concentración de ALT en dicho espécimen y desviación típica igual a 4 U/l (véase el Ejemplo 2.2.12). Suponga que el laboratorio de un cierto hospital mide muchos especímenes cada día, realizando un ensayo en cada espécimen, y que los especímenes con valores de ALT mayores o iguales que 40 U/l se clasifican como «inusualmente altos». Si la verdadera concentración de ALT de un paciente es de 35 U/l, ¿cuál es la probabilidad de que su espécimen se clasifique como «inusualmente alto»?

4.S.16 Se midió la frecuencia cardíaca en reposo de un grupo de sujetos. Después, los sujetos bebieron 6 onzas de café. 10 minutos más tarde se volvieron a medir sus frecuencias cardíacas. El cambio en la frecuencia cardíaca siguió una distribución normal, con un incremento medio de 7,3 latidos por minuto y una desviación típica de 11,1¹⁹. Sea Y la variable que indica el cambio en la frecuencia cardíaca de una persona seleccionada aleatoriamente. Calcule

- (a) $\Pr\{Y > 10\}$.
- (b) $\Pr\{Y > 20\}$.
- (c) $\Pr\{5 < Y < 15\}$.

4.S.17 En relación con la distribución de las frecuencias cardíacas del Ejercicio 4.S.16, el hecho de que la desviación típica sea mayor que la media y que la distribución sea normal nos indica que algunos de los valores de los datos son

negativos, lo que significa que en algunas personas la frecuencia cardíaca disminuyó en vez de aumentar. Calcule la probabilidad de que la frecuencia cardíaca de una persona elegida aleatoriamente disminuya. Es decir, calcule $\Pr\{Y < 0\}$.

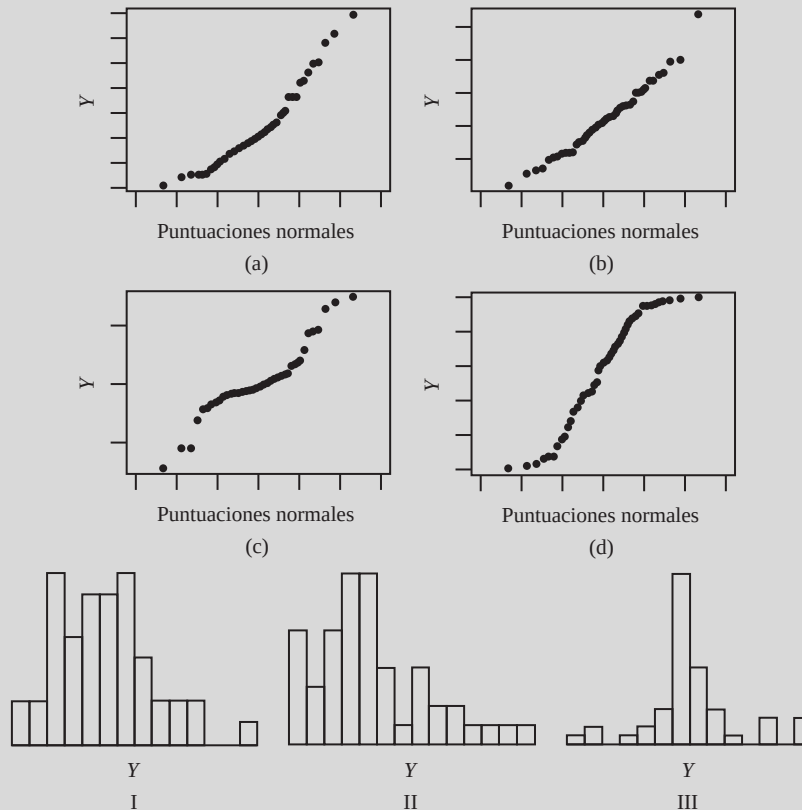
4.S.18 En relación con la distribución de las frecuencias cardíacas del Ejercicio 4.S.16, suponga que tomamos una muestra aleatoria de tamaño 400 de dicha distribución. ¿Cuántas observaciones esperaríamos obtener que estuvieran entre 0 y 15?

4.S.19 En relación con la distribución de las frecuencias cardíacas del Ejercicio 4.S.16, si utilizamos la regla de $1,5 \times \text{IQR}$, del Capítulo 2, para identificar *outliers*, ¿cómo de grande necesitaría ser una observación para ser considerada como *outlier* respecto al extremo superior?

4.S.20 Se ha dicho que las frecuencias cardíacas del Ejercicio 4.S.16 siguen una distribución normal. Si esto es cierto, ¿cuáles de los siguientes P -valores de la prueba de Shapiro-Wilk de una muestra aleatoria de 15 sujetos son consistentes con esta afirmación?

- (a) P -valor = 0,0149.
- (b) P -valor = 0,1345.
- (c) P -valor = 0,0498.
- (d) P -valor = 0,0042.

4.S.21 Las cuatro gráficas de probabilidad normal siguientes, (a), (b), (c) y (d), se han generado a partir de las distribuciones que se muestran en los histogramas I, II y III, y de otro histograma que no se muestra. ¿Qué gráfica de probabilidad normal corresponde a cada histograma? ¿Cómo lo sabe? (Habría una gráfica de probabilidad normal que no se utilice).



Notas

1. Datos del 2003-2004 National Health and Nutrition Examination Survey, que se pueden encontrar en www.denofinquiry.com/nhanes/source/choose.php
2. Ikeme, A. I., Roberts, C., Adams, R. L., Hester, P. Y. y Stadelman, W. J. (1983). Effects of supplementary water-administered vitamin D₃ on egg shell thickness. *Poultry Science* **62**, 1120-1122. La curva normal se ajustó a datos iniciales proporcionados por cortesía de W. J. Stadelman y A. I. Ikeme.
3. Hengstenberg, R. (1971). Das Augemuskelssystem der Stubenfliege *Musca domestica*. 1. Analyse der «clock-spikes» und ihrer Quellen. *Kybernetik* **2**, 56-57.
4. Adaptado de Magath, T. B. y Betkson, J. (1960). Electronic blood-cell counting. *American Journal of Clinical Pathology* **34**, 203-213. Realmente, el error porcentual es algo menor para cuentas altas y algo mayor para cuentas bajas. Descrito en Coulter Electronics (1982). *Performance Characteristics and Specifications for Coulter Counter Model S-560*. Hialeah, FL: Coulter Electronics.
5. Población ficticia pero realista. Adaptado de datos proporcionados por Hildebrand, S. F. y Schroeder, W. C. (1927). Fishes of Chesapeake Bay. *Bulletin of the United States Bureau of Fisheries* **43**, Part 1, p. 88. Los peces son jóvenes en ese año; observados en octubre, son muy pequeños. (La distribución de longitudes en poblaciones de más edad no es aproximadamente normal).
6. Adaptado de Pearl, R. (1905). Biometrical studies on man. I. Variation and correlation in brain weight. *Biometrika* **4**, 13-104.
7. Adaptado de Swearingen, M. L. y Halt, D. A. (1976). Using a «blank» trial as a teaching tool. *Journal of Agronomic Education* **5**, 3-8. La desviación típica dada en este problema es realista para un campo idealizado «uniforme», en el que las diferencias de cosechas entre parcelas se debieran a variación local aleatoria y no a variación a gran escala y quizá sistemática.
8. Adaptado de Coulter Electronics (1982). *Performance Characteristics and Specifications for the Coulter Counter Model S-560*. Hialeah, FL: Coulter Electronics.
9. Datos no publicados cortesía de Susan Whitehead, Oberlin College.
10. Datos tomados de www.athlinks.com/results/50228/97027/u1/2008-Rome-Marathon.aspx
11. Datos no publicados cortesía de Kaelyn Stiles, Oberlin College.
12. Datos no publicados cortesía de Paul Harnik y Lydia Ries, Oberlin College.
13. Resumen de información meteorológica obtenida de www.centralcoastweather.net
14. Resumen de información meteorológica obtenida de www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?akjune
15. Long, E. C. (1976). *Liquid Scintillation Counting Theory and Techniques*. Irvine, Calif.: Beckman Instruments. La distribución es realmente una distribución discreta denominada distribución de Poisson. Sin embargo, una distribución de Poisson de media grande es aproximadamente normal.
16. Población ficticia pero realista, basada en los datos de Emerson, R. A. y East, E. M. (1913). Inheritance of quantitative characters in maize. *Nebraska Experimental Station Research Bulletin* **2**. Datos reproducidos por Mather, K. (1943). *Statistical Analysis in Biology*. London: Methuen. pp. 29, 34. El maíz híbrido moderno es más alto y menos variable que el de esta población.
17. Estos percentiles se basan en datos del National Health and Nutrition Examination Survey (NHANES), realizado por el National Center for Health Statistics Centers for Disease Control and Prevention. La URL siguiente es un enlace a la tabla de datos: www.cdc.gov/nchs/about/major/nhanes/hgtfem.pdf
18. Esta es la distribución estándar de referencia para las puntuaciones de Stanford-Binet. Véase Sattler, J. M. (1982). *Assessment of Children's Intelligence and Special Abilities*, 2.^a ed. Boston: Allyn and Bacon, p. 19 y cubierta trasera.
19. Datos no publicados cortesía de Forrest Crawford e Yvonne Piper, Oberlin College.

DISTRIBUCIONES MUESTRALES

5

Objetivos

En este capítulo desarrollaremos la idea de distribución muestral, que es central de la inferencia estadística clásica. En particular

- describiremos las distribuciones muestrales;
- demostraremos como se relaciona el tamaño de la muestra con exactitud de la media muestral;
- exploraremos el Teorema Central del Límite.
- demostraremos cómo se puede utilizar la distribución normal para aproximar a la distribución binomial.

5.1 Ideas básicas

Un objetivo importante del análisis de datos es distinguir entre las características de los datos que reflejan hechos biológicos reales y las características que pueden reflejar solo efectos debidos al azar. Como se explicó en las Secciones 1.3 y 2.8, el modelo del muestreo aleatorio proporciona un marco para efectuar esta distinción. La realidad subyacente se visualiza como una población, los datos se ven como una muestra aleatoria de la población y los efectos del azar se consideran como error de muestreo: es decir, una discrepancia entre la muestra y la población.

En este capítulo desarrollaremos el marco teórico que nos permitirá poner límites específicos al grado de error de muestreo que se puede esperar en un estudio. (Aunque en el Capítulo 1 distinguimos entre estudio experimental y estudio observacional, en esta presentación denominaremos *estudio* a cualquier investigación científica). Como en capítulos anteriores, continuaremos restringiendo la presentación al contexto sencillo de un estudio con solo un grupo (una muestra).

Variabilidad del muestreo

La variabilidad de las muestras aleatorias de la misma población se denomina **variabilidad del muestreo**. Una distribución de probabilidad que caracteriza algún aspecto de la variabilidad del muestreo se denomina **distribución muestral**. En general, una muestra aleatoria tendrá aspectos similares a la población de la que procede. Por supuesto, tenemos que esperar un cierto grado de discrepancia entre la muestra de la población. Una distribución muestral nos indica la verosimilitud del grado de semejanza entre la muestra y la población.

En este capítulo presentaremos varios aspectos de la variabilidad del muestreo y estudiaremos una importante distribución muestral. A partir de este punto, asumiremos que el tamaño de la muestra es una fracción despreciable del tamaño de la población. Este supuesto simplifica la teoría, porque garantiza que el proceso de extraer la muestra no cambia la composición de la población de ninguna forma apreciable.

El metaestudio

De acuerdo con el modelo de muestreo aleatorio, consideramos los datos de un estudio como una muestra aleatoria de una población. Generalmente, solo obtenemos una única muestra aleatoria, que proviene de una población muy grande. Sin embargo, para visualizar la variabilidad del muestreo debemos ampliar nuestro marco de referencia para incluir no meramente una muestra, sino todas las posibles muestras que podrían haber sido extraídas de la población. Este marco de referencia más amplio se denominará **metaestudio**. Un metaestudio consiste en un número indefinido de repeticiones, o réplicas, del mismo estudio*. Por tanto, si el estudio consiste en extraer una muestra aleatoria de tamaño n de alguna población, el correspondiente metaestudio consiste en extraer *repetidamente* muestras aleatorias de tamaño n de la misma población. El proceso de extracción repetida se realiza indefinidamente, y los miembros de cada muestra son reemplazados antes de extraer la muestra siguiente. El estudio y el metaestudio se representan esquemáticamente en la Figura 5.1.1.

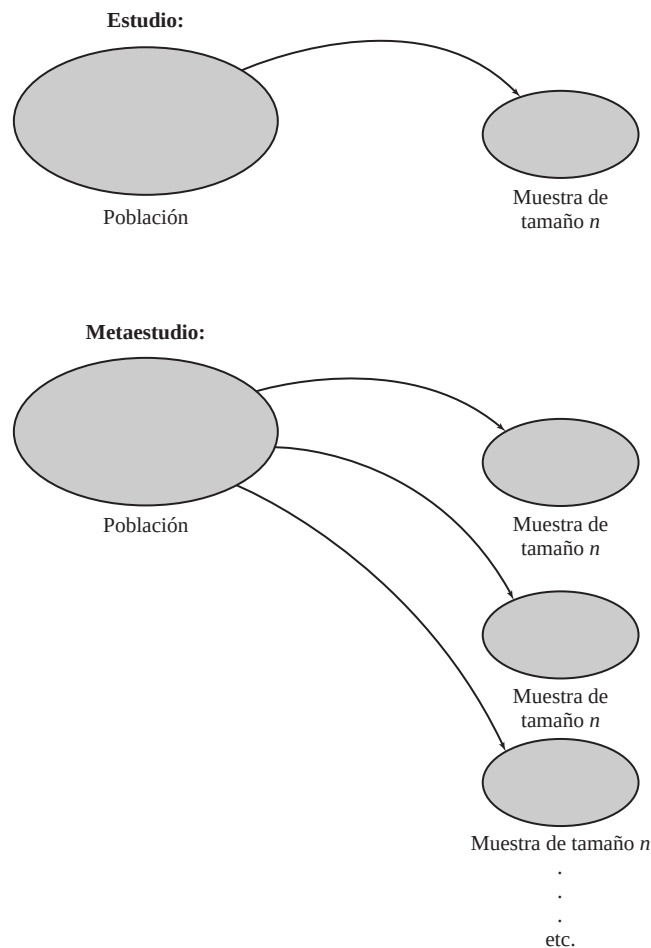


Figura 5.1.1 Representación esquemática de un estudio y un metaestudio

* El término *metaestudio* no es un término estándar. No está relacionado con el término *metaanálisis* que indica un tipo particular de análisis estadístico.

Ejemplo 5.1.1 Presión sanguínea en ratas

Un estudio consiste en medir el cambio en la presión sanguínea en cada una de $n = 10$ ratas tras administrarles un cierto medicamento. El correspondiente metaestudio consistiría en elegir repetidamente grupos de $n = 10$ ratas de la misma población y realizar las medidas de presión sanguínea bajo las mismas condiciones. ■

Ejemplo 5.1.2 Crecimiento bacteriano

Un estudio consistiría en observar el crecimiento bacteriano en $n = 5$ placas de Petri que se han tratado idénticamente. El correspondiente metaestudio consistiría en preparar repetidamente grupos de cinco placas de Petri y observarlas de la misma forma. ■

Nótese que un metaestudio es una construcción teórica en vez de una operación que un investigador realice normalmente.

El concepto de metaestudio proporciona una relación entre la variabilidad en el muestreo y la probabilidad. Recuerde del Capítulo 3 que la probabilidad de un suceso se puede interpretar como la frecuencia relativa asintótica de la ocurrencia de dicho evento. Escoger una muestra aleatoria es una operación aleatoria; el metaestudio consiste en muchas repeticiones de esta operación aleatoria y, por tanto, *las probabilidades concernientes a una muestra aleatoria se pueden interpretar como frecuencias relativas en un metaestudio*. Por consiguiente, el metaestudio es un instrumento para visualizar explícitamente una distribución muestral: la distribución muestral describe la variabilidad, para un estadístico dado, entre las muchas muestras aleatorias en un metaestudio.

Consideramos que ilustrar la idea de una distribución en el muestreo es un ejemplo pequeño (y artificial).

Ejemplo 5.1.3 Sustitución de rodilla

Considere una población de mujeres entre 65 y 75 años de edad que tienen problemas de rodilla y son candidatas a una cirugía de sustitución de rodilla. Una mujer podría recibir dicha cirugía de sustitución en una rodilla con un coste de 35.000 \$, en ambas rodillas con un coste de 60.000 \$ (una «sustitución doble», que es menos costosa que dos sustituciones por separado) o en ninguna rodilla. Consideremos la perspectiva de una compañía de seguros con respecto a una muestra de $n = 3$ mujeres que asegura. ¿Cuál es el coste total de tratar a esas tres mujeres? El valor más pequeño del coste total es cero (si ninguna de las tres mujeres necesita la cirugía), mientras que el máximo coste posible sería de 180.000 \$ (si las tres mujeres necesitaran una sustitución doble). Para hacer las cosas relativamente simples, supongamos que la cuarta parte de las mujeres entre 65 y 75 años requieren una sustitución doble, la mitad requiere la sustitución de una sola rodilla y una cuarta parte no requiere cirugía.

La lista completa de posibles muestras se presenta en la Tabla 5.1.1, junto con el total muestral (en miles de dólares) para cada caso y la probabilidad de que aparezca cada caso. Por ejemplo, la probabilidad de que las tres mujeres no requieran cirugía («Ninguna, Ninguna, Ninguna») es $(1/4) \times (1/4) \times (1/4) = 1/64$, mientras que la probabilidad de que las dos primeras mujeres no requieran cirugía y la tercera requiera una operación de una sola rodilla («Ninguna, Ninguna, Una») es $(1/4) \times (1/4) \times (2/4) = 2/64$. Hay 10 posibles valores del total muestral: 0, 35, 60, 70, 95, 105, 120, 130, 155 y 180. Las columnas primera y tercera de la Tabla 5.1.2 muestran las distribuciones muestrales del total muestral, combinando las muestras que producen el mismo total y sumando sus probabilidades. Por ejemplo, hay tres formas de que el total sea 70, y cada una tiene una probabilidad de $4/64$, por lo que su suma es $12/64$.

La segunda columna de la Tabla 5.1.2 muestra la media muestral (redondeada a un dígito decimal) de forma que las dos últimas columnas de la tabla presentan la distribución muestral de la media muestral. Estas dos distribuciones, que se muestran gráficamente en la Figura 5.1.2, se relacionan entre sí mediante un cambio de escala. Una compañía de seguros podría hablar en términos de coste total, pero esto es equivalente a considerar el coste promedio. ■

Tabla 5.1.1 Costes totales de sustitución de rodilla para todas las posibles muestras de tamaño $n = 3$			
Muestra	Costes (en unidades de 1.000 \$)	Total muestra	Probabilidad
Ninguna, Ninguna, Ninguna	0,0,0	0	1/64
Ninguna, Ninguna, Una	0,0,35	35	2/64
Ninguna, Ninguna, Dos	0,0,60	60	1/64
Ninguna, Una, Ninguna	0,35,0	35	2/64
Ninguna, Una, Una	0,35,35	70	4/64
Ninguna, Una, Dos	0,35,60	95	2/64
Ninguna, Dos, Ninguna	0,60,0	60	1/64
Ninguna, Dos, Una	0,60,35	95	2/64
Ninguna, Dos, Dos	0,60,60	120	1/64
Una, Ninguna, Ninguna	35,0,0	35	2/64
Una, Ninguna, Una	35,0,35	70	4/64
Una, Ninguna, Dos	35,0,60	95	2/64
Una, Una, Ninguna	35,35,0	70	4/64
Una, Una, Una	35,35,35	105	8/64
Una, Una, Dos	35,35,60	130	4/64
Una, Dos, Ninguna	35,60,0	95	2/64
Una, Dos, Una	35,60,35	130	4/64
Una, Dos, Dos	35,60,60	155	2/64
Dos, Ninguna, Ninguna	60,0,0	60	1/64
Dos, Ninguna, Una	60,0,35	95	2/64
Dos, Ninguna, Dos	60,0,60	120	1/64
Dos, Una, Ninguna	60,35,0	95	2/64
Dos, Una, Una	60,35,35	130	4/64
Dos, Una, Dos	60,35,60	155	2/64
Dos, Dos, Ninguna	60,60,0	120	1/64
Dos, Dos, Una	60,60,35	155	2/64
Dos, Dos, Dos	60,60,60	180	1/64

Relación con la inferencia estadística

Conocer una distribución muestral nos permite hacer afirmaciones de probabilidad sobre posibles muestras. Por ejemplo, en el contexto del Ejemplo 5.1.3, la compañía de seguros podría preguntar: ¿Cuál es la probabilidad de que los costes totales de sustitución de rodillas de una muestra de tres mujeres sean menores que 110.000 \$? Podemos responder a esta pregunta sumando las probabilidades de los seis primeros resultados que se muestran en la Tabla 5.1.2. La suma es 42/64. Abundaremos en esta idea cuando desarrollemos formalmente las ideas de la inferencia estadística.

Tabla 5.1.2 Distribución muestral de los costes totales de cirugía para muestras de tamaño $n = 3$

Total muestra	Media muestral	Probabilidad
0	0,0	1/64
35	11,7	6/64
60	20,0	3/64
70	23,3	12/64
95	31,7	12/64
105	35,0	8/64
120	40,0	3/64
130	43,3	12/64
155	51,7	6/64
180	60,0	1/64

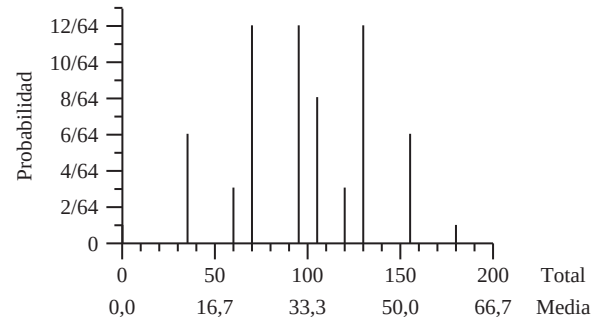


Figura 5.1.2 Gráfica de la distribución muestral de los costes totales de cirugía para muestras de tamaño $n = 3$

Ejercicios 5.1.1-5.1.4

5.1.1 Considere tomar una muestra aleatoria de tamaño tres de la población de sustitución de rodilla del Ejemplo 5.1.3. ¿Cuál es la probabilidad de que el coste total de los valores en la muestra sea mayor que 125.000 \$?

5.1.2 Considere tomar una muestra aleatoria de tamaño tres de la población de sustitución de rodilla del Ejemplo 5.1.3. ¿Cuál es la probabilidad de que el coste total de los valores en la muestra esté entre 80.000 \$ y 125.000 \$?

5.1.3 Considere tomar una muestra aleatoria de tamaño tres de la población de sustitución de rodilla del Ejemplo 5.1.3.

¿Cuál es la probabilidad de que el coste medio de los valores en la muestra esté entre 40.000 \$ y 100.000 \$?

5.1.4 Considere una población hipotética de perros en las que hay cuatro posibles pesos, todos igualmente probables: 42, 48, 52 o 58 libras. Si se toma una muestra de tamaño $n = 2$ de la población, ¿cuál es la distribución muestral del peso total de los dos perros seleccionados? Es decir, ¿cuáles son los posibles valores del total y cuáles son las probabilidades asociadas con cada uno de esos valores?

5.2 La media muestral

Dada una variable cuantitativa, la muestra y la población se pueden describir de varias formas: mediante la media, la mediana, la desviación típica, etc. Las naturalezas (es decir, forma, centro, dispersión) de las distribuciones muestrales de estas medidas descriptivas no son todas las mismas. En esta sección nos centraremos principalmente en la distribución muestral de la media.

La distribución muestral de $\bar{Y}$

La media muestral $\bar{y}$ se puede usar, no solo como la descripción de los datos de la muestra, sino también como un estimador de la media poblacional μ . Es natural preguntarse: «¿cuánto de cerca está $\bar{y}$ de μ ?». No podemos responder a esta pregunta para la media $\bar{y}$ de una muestra concreta, pero podemos responder si pensamos en términos del modelo de muestreo aleatorio y vemos la media muestral como una variable aleatoria $\bar{Y}$. La pregunta entonces se convierte en: «¿cuánto de cerca está $\bar{Y}$ de μ ?» y la respuesta la proporciona la **distribución muestral** de $\bar{Y}$, es decir, la distribución de probabilidad que describe la variabilidad en el muestreo de $\bar{Y}$.

Para visualizar la distribución muestral de $\bar{Y}$, imaginemos el metaestudio como sigue: se extraen repetidamente muestras aleatorias de tamaño n de una población fija con media μ y desviación típica σ . Cada muestra tiene su propia

media $\bar{y}$. La variación de los valores de $\bar{y}$ de una muestra a otra se especifica mediante la distribución muestral de $\bar{Y}$. Esta relación se indica esquemáticamente en la Figura 5.2.1.

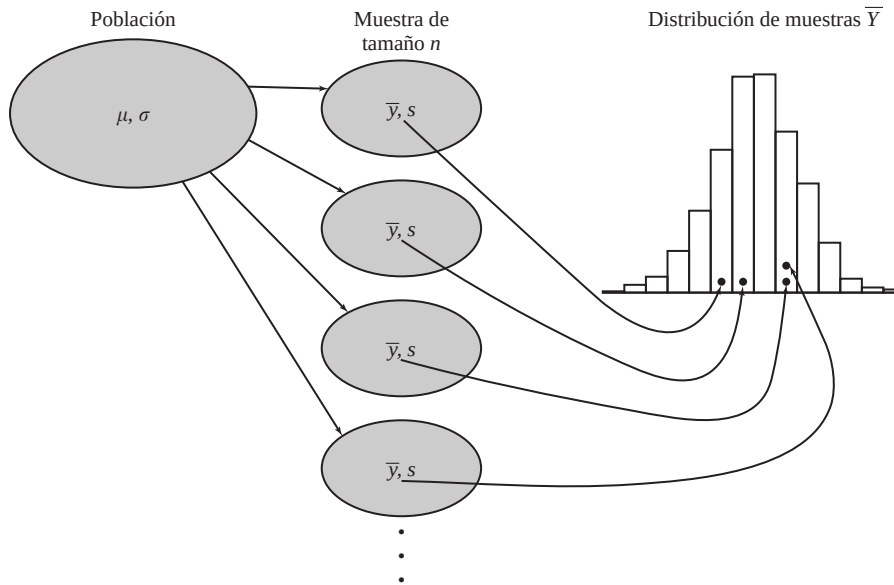


Figura 5.2.1 Representación esquemática de la distribución muestral de $\bar{Y}$

Cuando pensamos en $\bar{Y}$ como una variable aleatoria, necesitamos estar seguros de dos hechos básicos. El primero de ellos es intuitivo: en promedio, la media muestral será igual a la media poblacional. Es decir, el promedio de la distribución al muestreo de $\bar{Y}$ es μ . El segundo hecho no es obvio: la desviación típica de $\bar{Y}$ es igual a la desviación típica de Y dividida por la raíz cuadrada del tamaño muestral. Es decir, la desviación típica de $\bar{Y}$ es $\sigma/\sqrt{n}$.

Ejemplo 5.2.1 Colesterol en suero

Los niveles de colesterol en suero de niños entre 12 y 14 años siguen una distribución normal de media $\mu = 162$ mg/dl y desviación típica $\sigma = 28$ mg/dl¹. Si tomamos una muestra aleatoria, es razonable pensar que la media muestral esté próxima a 162, siendo las medias de algunas muestras superiores a 162 y de otras inferiores a 162. Como indica la fórmula anterior, el grado de variabilidad de la media muestral depende del grado de variabilidad de los niveles de colesterol de la población, σ . Si la población es muy homogénea (todos tienen aproximadamente el mismo valor de colesterol de forma que σ es pequeña), entonces las muestras y, por tanto, las medias muestrales serán todas muy similares y presentarán una variabilidad baja. Si la población es muy heterogénea (σ es grande), entonces las muestras (y por tanto los valores de la media muestral) variarán más. Aunque los investigadores tienen poco control sobre el valor de σ , podemos controlar el tamaño de la muestra, n , y n afecta a la cantidad de variabilidad de la media muestral. Si tomamos una muestra de tamaño $n = 9$, entonces la desviación típica de la media muestral es $\frac{28}{\sqrt{9}} = \frac{28}{3} = 9,3$. Esto significa, hablando en términos aproximados, que la media muestral, $\bar{Y}$, variará de una muestra a la siguiente aproximadamente 9,3 mg/dl*. Si tomáramos muestras aleatorias más grandes de tamaño $n = 25$, entonces la desviación típica de la media muestral sería más pequeña: $\frac{28}{\sqrt{25}} = \frac{28}{5} = 5,6$,

* Hablando en términos estrictos, la desviación típica mide la desviación con respecto a la media, no la diferencia entre observaciones consecutivas.

lo que quiere decir que $\bar{Y}$ variaría de una muestra a la siguiente aproximadamente 5,6. A medida que el tamaño muestral crece, la variabilidad de la media muestral $\bar{Y}$ se reduce. ■

Plantearemos ahora en forma de teorema los hechos básicos sobre la distribución muestral de $\bar{Y}$. Este teorema se puede demostrar utilizando los métodos de estadística matemática. Nosotros lo plantearemos sin demostración. El teorema describe la distribución muestral de $\bar{Y}$ en términos de su media (denominada $\mu_{\bar{Y}}$), su desviación típica (denominada $\sigma_{\bar{Y}}$) y su forma*.

Teorema 5.2.1: la distribución en el muestreo de $\bar{Y}$

1. **Media** La media de la distribución en el muestreo de $\bar{Y}$ es igual a la media poblacional. En símbolos,

$$\mu_{\bar{Y}} = \mu$$

2. **Desviación típica** La desviación típica de la distribución en el muestreo de $\bar{Y}$ es igual a la desviación típica poblacional dividida por la raíz cuadrada del tamaño de la muestra. En símbolos,

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

3. **Forma**

- (a) Si la distribución poblacional de Y es normal, entonces la distribución en el muestreo de $\bar{Y}$ es normal, independientemente del tamaño de la muestra n .
 (b) *Teorema Central del Límite*. Si n es grande, entonces la distribución en el muestreo de $\bar{Y}$ es aproximadamente normal, incluso aunque la distribución poblacional de Y no sea normal.

Los apartados 1 y 2 del Teorema 5.2.1 especifican la relación entre la media y la desviación típica de la población que está siendo muestreada y la media y la desviación típica de la distribución en el muestreo de $\bar{Y}$. El apartado 3(a) del teorema indica que, si la variable observada Y sigue una distribución normal en la población que está siendo muestreada, entonces la distribución en el muestreo de $\bar{Y}$ es también una distribución normal. Estas relaciones se indican en la Figura 5.2.2.

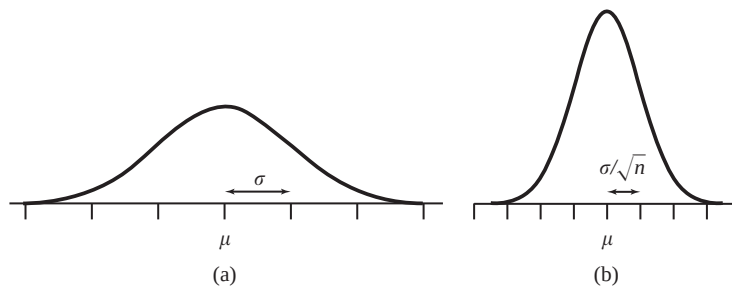


Figura 5.2.2 (a) La distribución poblacional de una variable Y distribuida normalmente; (b) la distribución en el muestreo de $\bar{Y}$ de las muestras de la población del apartado (a)

* Aquí estamos asumiendo que la población es infinitamente grande o, de forma equivalente, que estamos muestreando con reemplazamiento. Si muestreamos sin reemplazamiento de una población finita entonces se necesita un ajuste para obtener el valor correcto de

$\sigma_{\bar{Y}}$. En este caso $\sigma_{\bar{Y}}$ viene dada por $\frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$. El término $\sqrt{\frac{N-n}{N-1}}$ se denomina **factor de corrección de población finita**. Nótese

que si el tamaño de la muestra n es el 10 % del tamaño de la población N , entonces el factor de corrección es $\sqrt{\frac{0,9N}{N-1}} \approx 0,95$, de forma que el ajuste es pequeño. Por tanto, si n es pequeño en comparación con N , entonces el factor de corrección de población finita es cercano a 1 y se puede ignorar.

El siguiente ejemplo ilustra el significado de los apartados 1, 2 y 3(a) del Teorema 5.2.1.

Ejemplo 5.2.2 Pesos de semillas

Se va a muestrear una población grande de semillas de la judía *Phaseotus vulgaris*. Los pesos de las semillas de la población siguen una distribución normal de media $\mu = 500$ mg y desviación típica $\sigma = 120$ mg². Supongamos ahora que se va a pesar una muestra aleatoria de cuatro semillas, y sea $\bar{Y}$ la variable que representa el peso medio de las cuatro semillas. Entonces, de acuerdo con el Teorema 5.2.1, la distribución muestral de $\bar{Y}$ será una distribución normal con media y desviación típica como sigue:

$$\mu_{\bar{Y}} = \mu = 500 \text{ mg}$$

y

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{4}} = 60 \text{ mg}$$

Por tanto, en promedio, la media muestral será igual a 500 mg, pero la variabilidad de una muestra de tamaño 4 a la siguiente muestra de tamaño 4 es tal que aproximadamente las dos terceras partes de las veces $\bar{Y}$ estará dentro de un intervalo de 60 mg a cada lado alrededor de 500 mg, es decir, entre $500 - 60 = 440$ mg y $500 + 60 = 560$ mg. Asimismo, tomando dos desviaciones típicas, se puede esperar que $\bar{Y}$ esté dentro de un intervalo de 120 mg a cada lado de 500 mg, o entre $500 - 120 = 380$ mg y $500 + 120 = 620$ mg, aproximadamente el 95 % de las veces. La distribución muestral de $\bar{Y}$ se muestra en la Figura 5.2.3. Las marcas están separadas por una desviación típica. ■

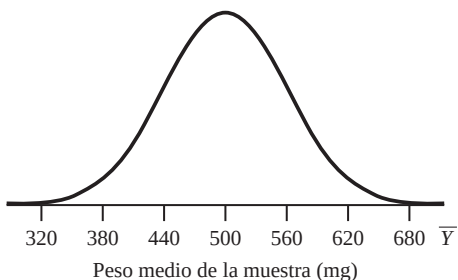


Figura 5.2.3 Distribución en el muestreo de $\bar{Y}$ en el Ejemplo 5.2.2

La distribución muestral de $\bar{Y}$ expresa la verosimilitud relativa de los diversos valores posibles de $\bar{Y}$. Un ejemplo, supongamos que deseamos saber la probabilidad de que el peso medio de las cuatro semillas sea mayor que 550 mg. Esta probabilidad se muestra como el área sombreada de la Figura 5.2.4. Nótese que el valor de $\bar{y} = 550$ se debe convertir a la escala Z utilizando la desviación típica $\sigma_{\bar{Y}} = 60$, y no $\sigma = 120$.

$$z = \frac{\bar{y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{550 - 500}{60} = 0,83$$

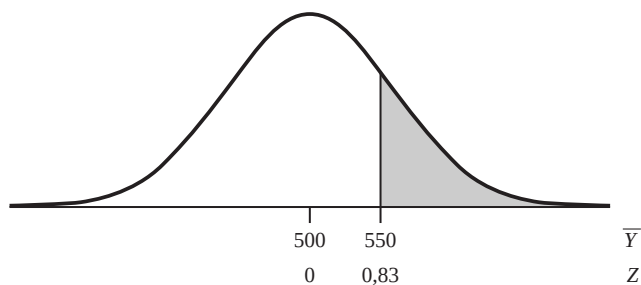


Figura 5.2.4 Cálculo de $\Pr\{\bar{Y} > 550\}$ en el Ejemplo 5.2.2

Utilizando la Tabla 3, $z = 0,83$ corresponde a un área de 0,7967. Por tanto,

$$\begin{aligned}\Pr\{\bar{Y} > 550\} &= \Pr\{Z > 0,83\} = 1 - 0,7967 \\ &= 0,2033 \approx 0,20\end{aligned}$$

Esta probabilidad se puede interpretar en términos de un metaestudio como sigue: si escogiéramos muchas muestras aleatorias de cuatro semillas cada una de la población, entonces aproximadamente el 20 % de las muestras tendría un peso medio superior a 550 mg.

El apartado 3(b) del Teorema 5.2.1 se conoce como **Teorema Central del Límite**. El Teorema Central del Límite establece que, *independientemente de qué distribución Y pueda haber en la población**, si el tamaño de la muestra es suficientemente grande, entonces la distribución muestral de $\bar{Y}$ es aproximadamente una distribución normal.

El Teorema Central del Límite es de importancia fundamental porque se puede aplicar cuando (como a menudo sucede en la práctica) la forma de la distribución poblacional no se conoce. Es debido al Teorema Central del Límite (y otros teoremas similares) que la distribución normal tiene un papel tan central en la estadística.

Es natural preguntarse cuánto de «grande» se requiere que sea el tamaño muestral para que se pueda aplicar el Teorema Central del Límite: ¿cuánto de grande debe ser n para que la distribución muestral de $\bar{Y}$ se aproxime bien mediante una curva normal? La respuesta es que el valor de n requerido depende de la forma de la distribución poblacional. Si la forma es normal, cualquier valor de n servirá. Si la forma es moderadamente no normal, un valor de n moderado será adecuado. Si la forma es altamente no normal, entonces se requerirá un valor de n más bien grande. (En la Sección 5.3 opcional se presentan algunos ejemplos concretos de este fenómeno).

Observación En la Sección 5.1 indicamos que la teoría de este capítulo es válida si el tamaño muestral es pequeño comparado con el tamaño de la población. Pero el Teorema Central del Límite es una afirmación sobre muestras grandes. Esto puede parecer una contradicción: ¿cómo puede una muestra grande ser una muestra pequeña? En la práctica, no hay contradicción. En una aplicación biológica típica, el tamaño de la población puede ser de 10^6 ; una muestra de tamaño $n = 100$ sería una pequeña fracción de la población pero, no obstante, sería lo suficientemente grande como para que el Teorema Central del Límite pudiera ser aplicable (en la mayoría de las situaciones).

Dependencia con el tamaño de la muestra

Consideremos la posibilidad de escoger muestras aleatorias de varios tamaños de la misma población. La distribución muestral de $\bar{Y}$ dependerá del tamaño de la muestra n de dos formas. En primer lugar, su desviación típica es

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

y es inversamente proporcional a $\sqrt{n}$. En segundo lugar, si la distribución poblacional no es normal, entonces la *forma* de la distribución en el muestreo de $\bar{Y}$ depende de n , y se aproxima más a la normal para n grande. Sin embargo, si la distribución poblacional es normal, entonces la distribución en el muestreo de $\bar{Y}$ es siempre normal, y solo depende de n la desviación típica.

El más importante de los dos efectos del tamaño de la muestra es el primero: valores de n grandes producen valores más pequeños de $\sigma_{\bar{Y}}$ y, en consecuencia, un error de muestreo esperado menor si $\bar{y}$ se utiliza como estimador de μ . El siguiente ejemplo ilustra este efecto para el muestreo de una población normal.

Ejemplo 5.2.3 Pesos de semillas

La Figura 5.2.5 muestra la distribución muestral de $\bar{Y}$ para muestras de varios tamaños de la población de judías del Ejemplo 5.2.2. Nótese que cuanto más grande es n , la distribución en el muestreo está más concentrada alrededor de la media poblacional $\mu = 500$ mg. Como consecuencia, la probabilidad de que $\bar{Y}$ esté cerca de dicha media poblacional es mayor cuanto más grande sea n . Por ejemplo, consideremos la probabilidad de que $\bar{Y}$ esté dentro de un

* Técnicamente, el Teorema Central del Límite requiere que la distribución de Y tenga una desviación típica. En la práctica, esta condición siempre se cumple.

intervalo de ± 50 mg alrededor de μ , es decir, $\Pr\{450 \leq \bar{Y} \leq 550\}$. La Tabla 5.2.1 muestra cómo depende esta probabilidad de n . ■

Tabla 5.2.1		
n	$\Pr\{450 \leq \bar{Y} \leq 550\}$	
4	0,59	
9	0,79	
16	0,91	
64	0,999	

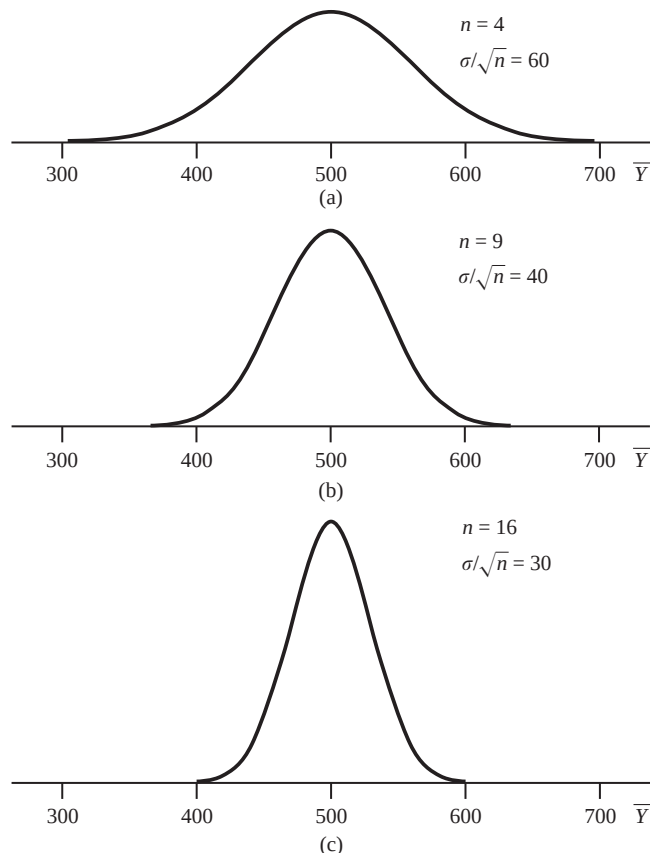


Figura 5.2.5 Distribución muestral de $\bar{Y}$ para varios tamaños de la muestra n

El Ejemplo 5.2.3 ilustra cómo depende del tamaño de la muestra la cercanía de $\bar{Y}$ a μ . La media de una muestra mayor no está *necesariamente* más cerca de la media poblacional que la media de una muestra menor, sino que tiene una *probabilidad mayor* de estar más cerca. Es en este sentido que una muestra mayor proporciona más información sobre la media poblacional que una muestra más pequeña.

Poblaciones, muestras y distribuciones muestrales

Al pensar sobre el Teorema 5.2.1, es importante distinguir claramente entre tres diferentes distribuciones relacionadas con una variable cuantitativa Y : (1) la distribución de Y en la población; (2) la distribución de Y en una muestra de datos, y (3) la distribución muestral de $\bar{Y}$. Las medias y las desviaciones típicas de estas distribuciones se resumen en la Tabla 5.2.2.

Tabla 5.2.2		
Distribución	Media	Desviación típica
Y en la población	μ	σ
Y en la muestra	$\bar{y}$	s
$\bar{Y}$ (en el metaestudio)	$\mu_{\bar{Y}} = \mu$	$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$

El ejemplo siguiente ilustra la distinción entre las tres distribuciones.

Ejemplo 5.2.4 Pesos de semillas

En la población de judías del Ejemplo 5.2.2, la media y la desviación típica poblacionales son $\mu = 500$ mg y $\sigma = 120$ mg. La distribución poblacional de $Y =$ peso se representa en la Figura 5.2.6(a). Supongamos que extraemos una muestra aleatoria de $n = 25$ semillas de la población y obtenemos los datos de la Tabla 5.2.3.

Para los datos de dicha tabla, la media muestral es $\bar{y} = 526,1$ mg y la desviación típica muestral es $s = 113,7$ mg. La Figura 5.2.6(b) muestra un histograma de los datos. Este histograma representa la distribución de Y en la muestra. La distribución muestral de $\bar{Y}$ es una distribución teórica que se relaciona, no con la muestra particular que se representa en el histograma, sino con el metaestudio de muestras repetidas de tamaño $n = 25$. La media y la desviación típica de la distribución en el muestreo son

$$\mu_{\bar{Y}} = 500 \text{ mg y } \sigma_{\bar{Y}} = \frac{120}{\sqrt{25}} = 24 \text{ mg}$$

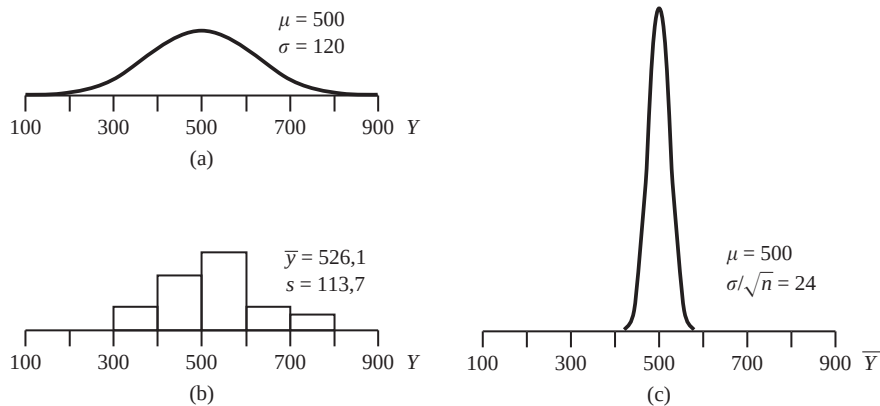


Figura 5.2.6 Tres distribuciones relacionadas con $Y =$ peso de semillas de judías: (a) distribución poblacional de Y ; (b) distribución de 25 observaciones de Y ; (c) distribución muestral de $\bar{Y}$ para $n = 25$.

Tabla 5.2.3 Pesos de 25 semillas de judías						
Peso (mg)						
343	755	431	480	516	469	694
659	441	562	597	502	612	549
348	469	545	728	416	536	581
433	583	570	334			

La distribución muestral se representa en la Figura 5.2.6(c). Nótese que las distribuciones en las Figuras 5.2.6(a) y (b) son más o menos similares; de hecho, la distribución en (b) es un estimador (basado en los datos de la Tabla 5.2.3) de la distribución en (a). Por el contrario, la distribución en (c) es mucho más estrecha, ya que que representa la distribución de *medias* en vez de observaciones individuales. ■

Otros aspectos de la variabilidad en el muestreo

La presentación anterior se ha enfocado en la variabilidad en el muestreo de la media muestral, $\bar{Y}$. Otros dos aspectos importantes de la variabilidad en el muestreo son (1) la variabilidad en el muestreo de la desviación típica muestral, s

y (2) la variabilidad en el muestreo de la *forma* de la muestra, representada por el histograma de dicha muestra. En vez de presentar formalmente estos aspectos, los ilustraremos con el ejemplo siguiente.

Ejemplo 5.2.5 Pesos de semillas

En la Figura 5.2.6(b) consideramos una muestra aleatoria de 25 observaciones de la población de judías del Ejemplo 5.2.2. Mostraremos ahora en la Figura 5.2.7 ocho muestras aleatorias adicionales de la misma población. (Las nueve muestras fueron realmente simuladas utilizando un computador). Nótese que, incluso aunque las muestras se han extraído de una población normal [que se muestra en la Figura 5.2.6(a)], existe una variación muy sustancial en las formas de los histogramas. Nótese también que existe una variación considerable en las desviaciones típicas muestrales. Por supuesto, si el tamaño de la muestra fuera mayor (por ejemplo, $n = 100$ en lugar de $n = 25$), habría menos variación en el muestreo. Los histogramas tenderían a parecerse más estrechamente a una curva normal y las desviaciones típicas tenderían a estar más cerca del valor poblacional ($\sigma = 120$). ■

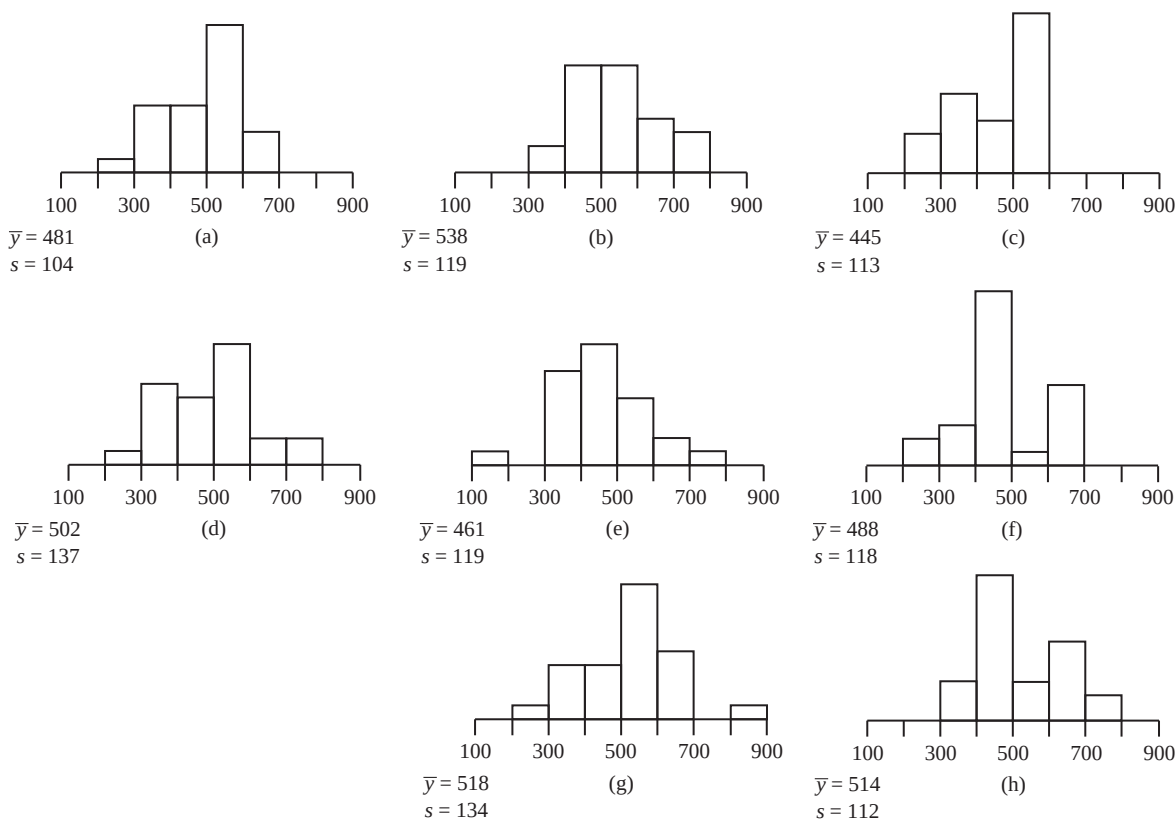


Figura 5.2.7 Ocho muestras aleatorias, todas de tamaño $n = 25$, de una población normal con $\mu = 500$ y $\sigma = 120$.

Ejercicios 5.2.1-5.2.19

5.2.1 (Ejercicio de muestreo) En relación con el Ejercicio 1.3.5, la colección de 100 elipses que se muestra en dicho ejemplo puede verse como una representación de una población natural del organismo *C. ellipticus*. Utilice su juicio para escoger una muestra de cinco elipses en que le parezcan

razonablemente representativas de la población. (Para simular mejor el juicio análogo en una situación de la vida real, debería realizar su selección de forma intuitiva, sin ningún estudio preliminar detallado de la población). Con una regla, mida longitud de cada elipse de su muestra. Mida sobre el

cuerpo, excluyendo las cerdas de la cola. Considere adecuadas medidas con una precisión de milímetro. Calcule la media y la desviación típica de las cinco longitudes. Para facilitar la recogida de resultados de toda la clase, exprese la media y la desviación típica en milímetros, con dos cifras decimales.

5.2.2 (Ejercicio de muestreo) Proceda como en el Ejercicio 5.2.1, pero utilice muestreo aleatorio en vez de muestreo «basado en su juicio». Para ello, escoja 10 dígitos aleatorios (de la Tabla 1 o utilizando su calculadora). Sean los primeros dos dígitos el número de la primera elipse que seleccionará en su muestra, y así sucesivamente. Los 10 dígitos aleatorios le permitirán extraer una muestra aleatoria de cinco elipses.

5.2.3 (Ejercicio de muestreo) Proceda como en el Ejercicio 5.2.1, pero escoja una muestra aleatoria de 20 elipses.

5.2.4 En relación con el Ejercicio 5.2.2, se propone el esquema siguiente para escoger una muestra de cinco elipses de la población de 100 elipses. (i) Escoja aleatoriamente un punto en el «hábitat» de las elipses (es decir, la figura); esto se puede hacer directamente lanzando un lápiz de punta sobre la página, o mucho mejor superponiendo en la página un papel milimetrado y utilizando dígitos aleatorios. (ii) Si el punto elegido está dentro de una elipse, incluya dicha elipse en la muestra; en otro caso comience de nuevo en el paso (i). (iii) Continúe hasta haber seleccionado 5 elipses. Explique por qué este esquema no es equivalente al muestreo aleatorio. ¿En qué dirección está sesgado el esquema, es decir, tendería a producir un valor de $\bar{y}$ demasiado grande a un valor de $\bar{y}$ demasiado pequeño?

5.2.5 Los niveles de colesterol en suero de una población de niños entre 12 y 14 años siguen una distribución normal de media 162 mg/dl y desviación típica 28 mg/dl (como en el Ejemplo 4.1.1).

- ¿Qué porcentaje de los niños entre 12 y 14 años tienen valores de colesterol en el suero entre 152 y 172 mg/dl?
- Suponga que deseáramos escoger aleatoriamente de la población un gran número de grupos de nueve niños entre 12 y 14 años cada uno. ¿En qué porcentaje de los grupos estaría el valor medio de grupo del colesterol entre 152 y 172 mg/dl?
- Si $\bar{Y}$ representa el valor medio de colesterol de una muestra aleatoria de nueve niños entre 12 y 14 años de la población, ¿cuánto vale $\Pr\{152 \leq \bar{Y} \leq 172\}$?

5.2.6 Un indicador importante de la función pulmonar es el volumen expiratorio forzado (VEF), que es el volumen de aire que una persona puede expirar en un segundo. El Dr. Hernández planea medir el VEF en una muestra aleatoria de n mujeres jóvenes de una cierta población, y utilizar la media muestral $\bar{y}$ como un estimador de la media poblacional.

Sea E el suceso de que la media muestral del Dr. Hernández esté dentro de un intervalo de ± 100 ml alrededor de la media poblacional. Suponga que la distribución poblacional es normal de media 3.000 ml y desviación típica 400 ml³. Calcule si

- $n = 15$.
- $n = 60$.
- ¿Cómo depende $\Pr\{E\}$ del tamaño de la muestra? Es decir, indique si cuando n crece, $\Pr\{E\}$ crece, decrece o no varía.

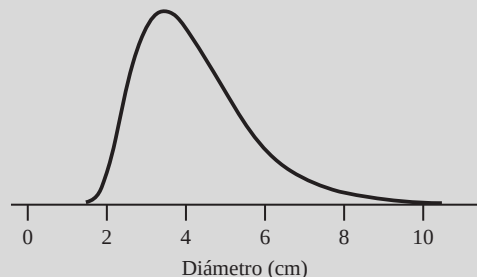
5.2.7 En relación con el Ejercicio 5.2.6, suponga que la distribución poblacional de VEF es normal con desviación típica 400 ml.

- Calcule $\Pr\{E\}$ si $n = 15$ si la media poblacional es de 2.800 ml.
- Calcule $\Pr\{E\}$ si $n = 15$ si la media poblacional es de 2.600 ml.
- ¿Cómo depende $\Pr\{E\}$ de la media poblacional?

5.2.8 Las alturas de una cierta población de plantas de maíz siguen una distribución normal de media 145 cm y desviación típica 22 cm (como en el Ejercicio 4.S.4).

- ¿Qué porcentaje de las plantas tiene una altura entre 135 y 155 cm?
- Suponga que deseáramos escoger aleatoriamente de la población un gran número de muestras de 16 plantas cada una. ¿En qué porcentaje de las muestras estaría la altura media muestral entre 135 y 155 cm?
- Si $\bar{Y}$ representa la altura media de una muestra aleatoria de 16 plantas de la población, ¿cuánto vale $\Pr\{135 \leq \bar{Y} \leq 155\}$?
- Si $\bar{Y}$ representa la altura media de una muestra aleatoria de 36 plantas de la población, ¿cuánto vale $\Pr\{135 \leq \bar{Y} \leq 155\}$?

5.2.9 El diámetro basal de una anémona de mar es un indicador de su edad. La curva de densidad que se muestra a continuación representa la distribución de diámetros de una cierta población grande de anémonas. El diámetro medio poblacional es de 4,2 cm, y la desviación típica es de 1,4 cm⁴. Sea $\bar{Y}$ la variable que representa el diámetro medio de 25 anémonas seleccionadas aleatoriamente de la población.



- (a) Calcule el valor aproximado de $\Pr\{4 \leq \bar{Y} \leq 5\}$.
 (b) ¿Por qué es su respuesta al apartado (a) aproximadamente correcta incluso aunque la distribución poblacional de los diámetros es claramente no normal? ¿Sería el planteamiento igualmente válido para una muestra de tamaño 2 en vez de 25? ¿Por qué o por qué no?

5.2.10 En una cierta población de peces, las longitudes de dichos peces siguen una distribución aproximadamente normal de media 54,0 mm y desviación típica 4,5 mm. Vimos en el Ejemplo 4.3.1 que en esta situación el 65,68 % de los peces miden entre 51 y 60 mm. Suponga que se extrae una muestra aleatoria de cuatro peces de la población. Calcule la probabilidad de que:

- (a) Todos los peces midan entre 51 y 60 mm.
 (b) La longitud media de los cuatro peces esté entre 51 y 60 mm.

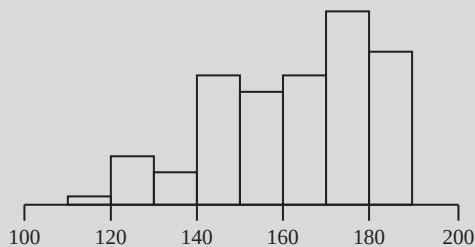
5.2.11 En el Ejercicio 5.2.10, la respuesta al apartado (b) fue mayor que la respuesta al apartado (a). Argumente que esto debe ser necesariamente cierto, independientemente de los valores de la media y desviación típica poblacionales. [Sugerencia: ¿podría suceder que el suceso del apartado (a) ocurriera pero que el suceso del apartado (b) no ocurriera?].

5.2.12 El profesor Smith realizó un ejercicio en clase en el que los estudiantes ejecutaron un programa de ordenador para generar muestras aleatorias de una población que tenía una media de 50 y una desviación típica de 9 mm. Cada uno de los estudiantes de Smith tomó una muestra aleatoria de tamaño n y calculó la media muestral. Smith encontró que aproximadamente el 68 % de los estudiantes habían calculado medias muestrales entre 48,5 y 51,5 mm. ¿Cuánto valía n ? (Suponga que n es lo suficientemente grande como para que sea aplicable el Teorema Central del Límite).

5.2.13 Un cierto ensayo sobre la alanina aminotransferasa (ALT) es algo impreciso. Los resultados de ensayos repetidos sobre un mismo espécimen siguen una distribución normal de media igual a la concentración de ALT en dicho espécimen y desviación típica igual a 4 U/l (como en el Ejercicio 4.S.15). Suponga que un laboratorio de un hospital mide muchos especímenes cada día, y aquellos especímenes cuyos valores de ALT son de 40 o más se etiqueta como «inusualmente altos». Si la verdadera concentración de ALT de un paciente es de 35 U/l, calcule la probabilidad de que su espécimen se ha etiquetado como «inusualmente alto»:

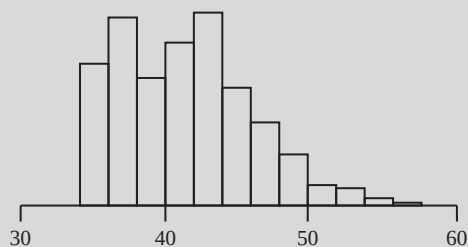
- (a) Si el valor indicado es el resultado de un único ensayo.
 (b) Si el valor indicado es la media de tres ensayos independientes sobre el mismo espécimen.

5.2.14 La media de la distribución que se muestra en el histograma siguiente es 162 y la desviación típica es 18. Considere que se toman muestras aleatorias de tamaño $n = 9$ de esta distribución y se calcula la media muestral, $\bar{y}$, de cada muestra.



- (a) ¿Cuál es la media de la distribución en el muestreo de $\bar{Y}$?
 (b) ¿Cuál es la desviación típica de la distribución en el muestreo de $\bar{Y}$?

5.2.15 La media de la distribución que se muestra en el histograma siguiente es 41,5 y la desviación típica es 4,7. Considere que se toman muestras aleatorias de tamaño $n = 4$ de esta distribución y se calcula la media muestral, $\bar{y}$, de cada muestra.



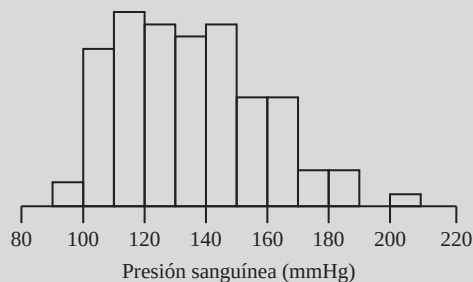
- (a) ¿Cuál es la media de la distribución muestral de $\bar{Y}$?
 (b) ¿Cuál es la desviación típica de la distribución en el muestreo de $\bar{Y}$?

5.2.16 En relación con el histograma del Ejercicio 5.2.15, suponga que se extraen 100 muestras aleatorias de esta población y se calcula la media muestral de cada muestra. Si se hiciera un histograma de la distribución de la medias muestrales de las 100 muestras, indique qué forma se podría esperar que tuviera el histograma:

- (a) Si $n = 2$ en cada muestra aleatoria.
 (b) Si $n = 25$ en cada muestra aleatoria.

5.2.17 En relación con el histograma del Ejercicio 5.2.15, suponga que se extraen 100 muestras aleatorias de esta población y se calcula la media muestral de cada muestra. Si se hiciera un histograma de la distribución de la medias muestrales de las 100 muestras, indique qué forma se podría esperar que tuviera el histograma si $n = 1$ para cada muestra aleatoria. Es decir, ¿qué aspecto tiene la distribución en el muestreo de la media cuando el tamaño de la muestra es $n = 1$?

5.2.18 Un investigador médico midió la presión sanguínea sistólica en 100 hombres de mediana edad⁵. Los resultados se muestran en el histograma que acompaña a este ejercicio. Nótese que la distribución está bastante sesgada. De acuerdo



con el Teorema Central del Límite, ¿podría esperarse que la distribución de las lecturas de presión sanguínea estuviera menos sesgada (y con una mayor forma de campana) si estuvieran basadas en $n = 400$ en vez de en $n = 100$ hombres? Explique su respuesta.

5.2.19 La presión parcial de oxígeno, PaO_2 , es una medida de la cantidad de oxígeno en la sangre⁶. Suponga que la distribución de los niveles de PaO_2 en recién nacidos tiene una media de 38 (mmHg) y la desviación típica de 9. Si se toma una muestra de tamaño $n = 25$:

- ¿Cuál es la probabilidad de que la media muestral sea mayor que 36?
- ¿Cuál es la probabilidad de que la media muestral sea mayor que 41?

5.3 Ilustración del Teorema Central del Límite (opcional)

La importancia de la distribución normal en estadística es debida en gran parte al Teorema Central del Límite y sus teoremas relacionados. En esta sección estudiaremos más de cerca el Teorema Central del Límite. De acuerdo con el Teorema Central del Límite, la distribución muestral de $\bar{Y}$ es aproximadamente normal si n es grande. Si consideramos muestras cada vez mayores de una población fija no normal, entonces la distribución muestral de $\bar{Y}$ se aproximará cada vez más a una normal para valores de n cada vez mayores. El siguiente ejemplo muestra cómo funciona el Teorema Central del Límite para dos distribuciones no normales: una distribución moderadamente sesgada (Ejemplo 5.3.1) y una distribución altamente sesgada (Ejemplo 5.3.2).

Ejemplo 5.3.1 Facetas de ojos

El número de facetas en el ojo de la mosca de la fruta *Drosophila melanogaster* es de interés en estudios genéticos. La distribución de esta variable en una cierta población de *Drosophila* se puede aproximar mediante la función de densidad que se muestra en la Figura 5.3.1. La distribución está moderadamente sesgada; la media y la desviación típica poblacionales son $\mu = 64$ y $\sigma = 22$ (ver notas⁷).

La Figura 5.3.2 muestra la distribución muestral de para muestras de varios tamaños de la población de facetas de ojos. Para mostrar claramente la forma de estas distribuciones, las hemos dibujado en diferentes escalas; la escala horizontal está más comprimida a medida que n es mayor. Nótese que las distribuciones están algo sesgadas hacia la derecha, pero el sesgo disminuye para valores de n mayores; para $n = 32$ la distribución se parece mucho a una normal. ■

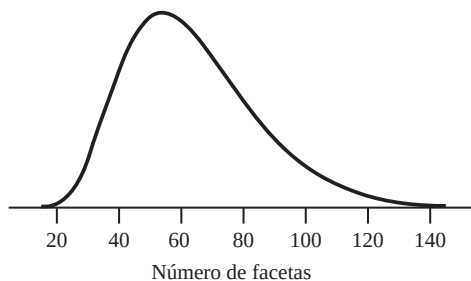


Figura 5.3.1 Distribución del número de facetas de ojos en una población de *Drosophila*

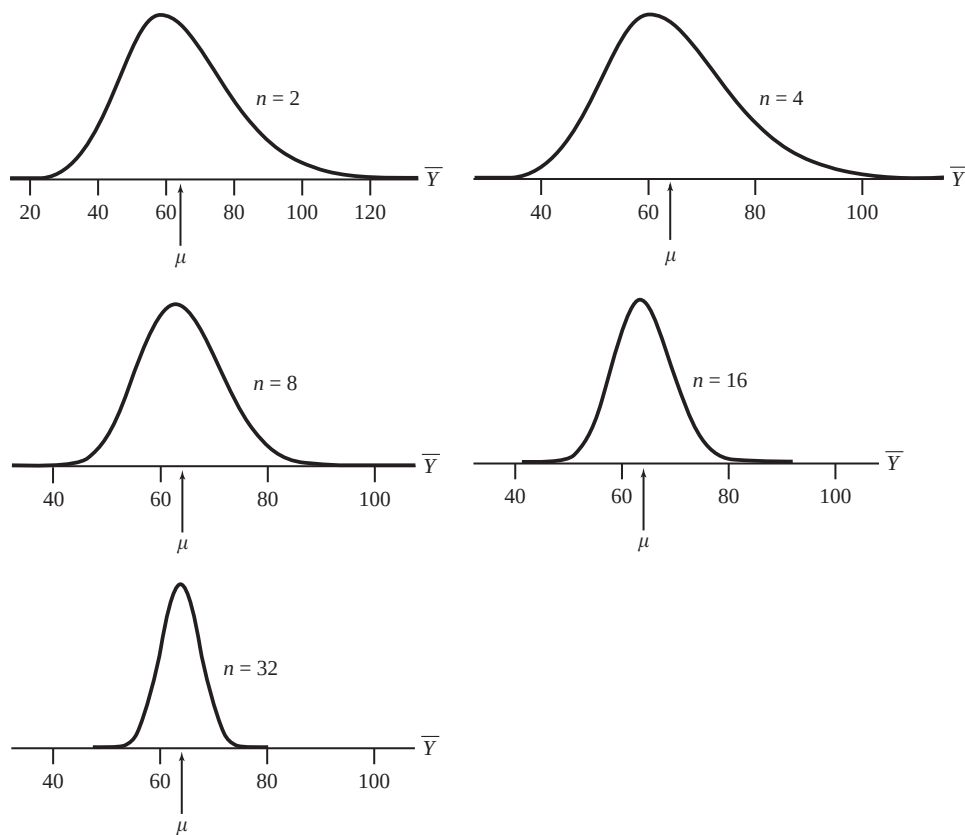


Figura 5.3.2 Distribuciones en el muestreo de $\bar{Y}$ para las muestras de la población de facetas de ojos de *Drosophila*

Ejemplo 5.3.2 Tiempo de reacción

Un psicólogo midió el tiempo requerido por una persona para levantar el brazo desde una posición fija y pulsar un botón con su dedo índice. La distribución de los tiempos (en milisegundos) de una única persona está representada por la densidad que se muestra en la Figura 5.3.3. Aproximadamente el 10 % del tiempo, el sujeto buscó a tientas o falló al pulsar el botón la primera vez. Los tiempos de retraso resultantes aparecen como segundo pico de la distribución⁸. El primer pico está centrado en 115 ms y el segundo en 450 ms. Debido a la presencia de los dos picos, la distribución global está altamente sesgada. La media y desviación típica poblacionales son $\mu = 148$ ms y $\sigma = 105$ ms, respectivamente.

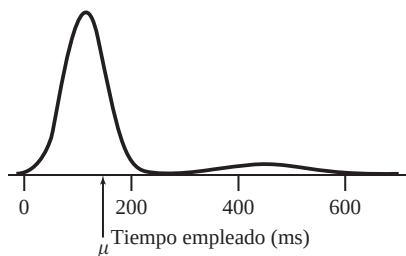


Figura 5.3.3 Distribución de tiempos en la prueba de pulsar un botón

La Figura 5.3.4 muestra la distribución muestral de $\bar{Y}$ para muestras de varios tamaños de la distribución de tiempos empleados. Para mostrar la forma claramente, la escala Y está más comprimida a medida que n es mayor. Nótese que para n pequeño la distribución tiene varias modas. A medida que n crece, las modas se reducen a pequeños bultos sobre la función de densidad y finalmente desaparecen, y la distribución se va haciendo cada vez más simétrica. ■

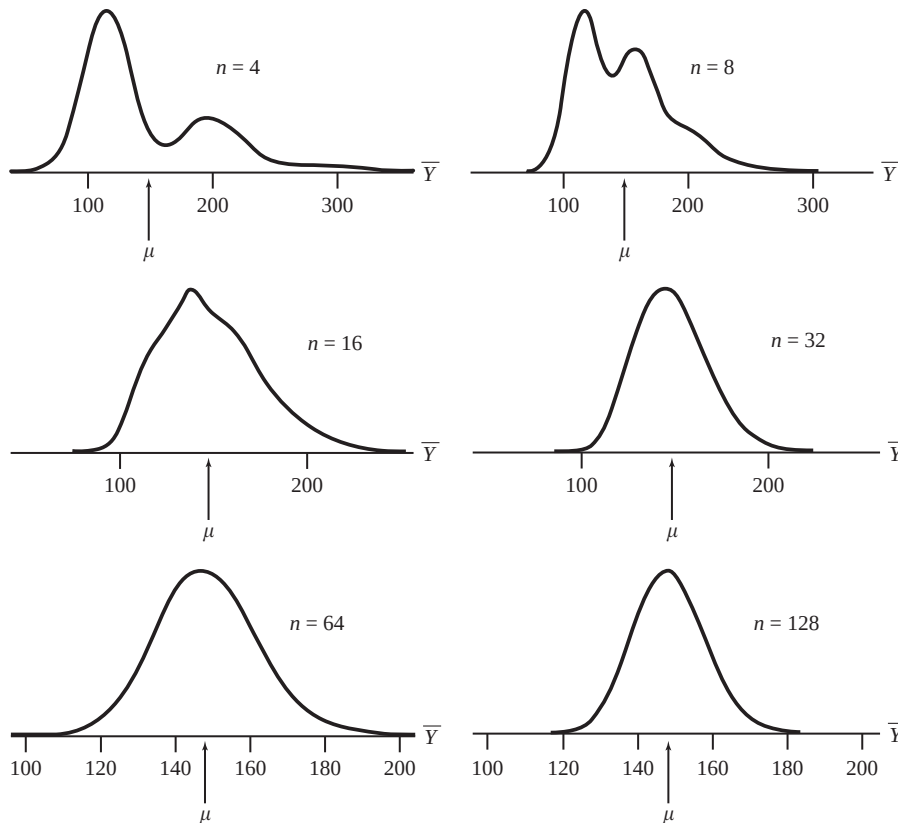


Figura 5.3.4 Distribución muestral de $\bar{Y}$ para las muestras de la población de tiempos empleados

Los Ejemplos 5.3.1 y 5.3.2 ilustran el hecho, mencionado la Sección 5.2, de que el significado del requisito « n es grande» en el Teorema Central del Límite depende de la forma de la distribución poblacional. La normalidad aproximada de la distribución muestral de $\bar{Y}$ se conseguirá para valores de n moderados si la distribución poblacional solo es moderadamente no normal (como en el Ejemplo 5.3.1), mientras que una población altamente no normal (como en el Ejemplo 5.3.2) requerirá valores de n mayores.

Nótese, sin embargo, que el Ejemplo 5.3.2 indica la notable fuerza del Teorema Central del Límite. El sesgo de la distribución de tiempos empleados es tan extremo que se podría ser reacio a considerar la media como una medida resumen. Incluso en este «caso peor», se puede ver el efecto del Teorema Central del Límite en la suavidad relativa y la simetría de la distribución muestral para $n = 64$.

El Teorema Central del Límite puede parecer como mágico. Para desmitificarlo de alguna forma, consideremos con más detalle en el ejemplo siguiente las distribuciones muestrales de los tiempos empleados.

Ejemplo 5.3.3 Tiempo de reacción

Consideremos las distribuciones muestrales de $\bar{Y}$ que se muestran en la Figura 5.3.4. Analicemos primero la distribución para $n = 4$, que es la distribución de la media de cuatro pulsaciones del botón. El pico alto a la izquierda

de la distribución representa los casos en el que el sujeto no operó a tientas ninguna de las 4 veces, de forma que los cuatro tiempos estuvieron alrededor de 115 ms; este resultado ocurriría el 66 % de las veces [a partir de la distribución binomial, porque $(0,9)^4 = 0,66$]. El pico más bajo siguiente representa los casos en los cuales en tres veces se emplearon aproximadamente 115 ms, y una vez se operó a tientas y se empleó aproximadamente 450 ms. (Nótese que el promedio de tres valores de 115 y uno de 450 es aproximadamente 200, que es el centro del segundo pico). De forma similar, el tercer pico (que apenas es visible) representa los casos en los que el sujeto operó a tientas 2 de las 4 veces. Los picos que representan las veces que el sujeto operó a tientas 3 y 4 veces son demasiado bajos para ser visibles en el dibujo.

Consideremos ahora la gráfica para $n = 8$. El primer pico representa 8 intentos buenos (sin operar a tientas), el segundo representa 7 intentos buenos y 1 en el que se operó a tientas, el tercero representa 6 intentos buenos y 2 en los que se operó a tientas, y así sucesivamente. Los picos cuarto y posteriores no se aprecian. Para $n = 16$ es más probable ver 15 intentos buenos y 1 en el que superó a tientas que 16 intentos buenos (como se puede verificar mediante la distribución binomial) y, por tanto, hay un abultamiento, correspondiente a 16 intentos buenos, por debajo del pico global, que corresponde a 15 intentos buenos; el abultamiento a la derecha del pico corresponde a 14 intentos buenos y 2 en los que se operó a tientas. Para $n = 32$, el resultado más probable es 3 operaciones a tientas y 29 intentos buenos; este resultado produce un tiempo medio de aproximadamente

$$\frac{(3)(450) + (29)(115)}{32} \approx 146 \text{ ms}$$

que es la posición del pico central. Por razones similares, las distribuciones para valores de n mayores están centradas aproximadamente en 148 ms, que es la media poblacional. ■

Ejercicios 5.3.1-5.3.3

5.3.1 En relación con el Ejemplo 5.3, en la distribución muestral de $\bar{Y}$ para $n = 4$ (Figura 5.3.4), indique cuál es el área aproximada bajo:

- (a) El primer pico.
- (b) El segundo pico

(Sugerencia: utilice la distribución binomial).

5.3.2 En relación con el Ejemplo 5.3.3, considere la distribución muestral de $\bar{Y}$ para $n = 2$ (que no se muestra en la Figura 5.3.4).

(a) Realice un dibujo aproximado de la distribución muestral. ¿Cuántos picos tiene? Muestre la posición (en el eje Y) de cada pico.

(b) Calcule el área aproximada debajo de cada pico. (Sugerencia: utilice la distribución binomial).

5.3.3 En relación con el Ejemplo 5.3.3, considere la distribución muestral de $\bar{Y}$ para $n = 2$ (que no se muestra en la Figura 5.3.4). Realice un dibujo aproximado de la distribución muestral. ¿Cuántos picos tiene? Muestre la posición (en el eje Y) de cada pico.

5.4 Aproximación normal a la distribución binomial (opcional)

El Teorema Central del Límite nos indica que la distribución muestral de la media adquiere forma de campana a medida que el tamaño de la muestra aumenta. Supongamos que tenemos una población dicotómica grande en la que etiquetamos los dos tipos de resultados como «1» (para el «éxito») y «0» (para el «fracaso»). Si tomamos una muestra y calculamos el número medio de unos, entonces esta media muestral es simplemente la proporción de unos (denominada comúnmente $\hat{P}$) y obedece al Teorema Central del Límite. Esto significa que si el tamaño de la muestra n es grande, entonces la distribución de $\hat{P}$ será aproximadamente normal.

Nótese que si sabemos el número de unos (es decir, el número de éxitos en n pruebas), entonces conocemos la proporción de unos, y viceversa. Por tanto, la aproximación normal a la distribución binomial se puede expresar de dos formas equivalentes: en términos del número de éxitos, Y , o en términos de la proporción de éxitos, $\hat{P}$. En el siguiente teorema plantaremos ambas formas. En este teorema, n representa el tamaño de la muestra (o, de forma más general, el número de muestras independientes) y p representa la proporción poblacional (o, de forma más general, la probabilidad de éxito en cada prueba independiente).

Teorema 5.4.1: aproximación normal a la distribución binomial

- (a) Si n es grande, entonces la distribución binomial del número de éxitos, Y , se puede aproximar mediante una distribución normal con

$$\text{Media} = np$$

y

$$\text{Desviación típica} = \sqrt{np(1-p)}$$

- (b) Si n es grande, entonces la distribución en el muestreo de $\hat{P}$ se puede aproximar mediante una distribución normal con

$$\text{Media} = p$$

y

$$\text{Desviación típica} = \sqrt{\frac{p(1-p)}{n}}$$

Observaciones

1. El Apéndice 5.1 proporciona una explicación más detallada de la relación entre la aproximación normal a la binomial y el Teorema Central del Límite.
2. Como se demuestra en el Apéndice 3.2, dada una población de ceros y unos donde la proporción de unos está dada por p , la desviación típica es $\sigma = \sqrt{p(1-p)}$. En el Teorema 5.2.1 se planteó que la desviación típica de una media está dada por $\frac{\sigma}{\sqrt{n}}$. Si vemos a $\hat{P}$ en el apartado (b) del Teorema 5.2.1 como una clase especial de media muestral, para el caso en el que todos los datos sean ceros y unos, el Teorema 5.2.1 indica que la desviación típica de $\hat{P}$ debería ser $\frac{\sqrt{p(1-p)}}{\sqrt{n}}$, o $\sqrt{\frac{p(1-p)}{n}}$, lo que está de acuerdo con el resultado planteado en el Teorema 5.4.1(b).

El siguiente ejemplo ilustra el uso del Teorema 5.4.1.

Ejemplo 5.4.1 Aproximación normal a la binomial

Consideremos una distribución binomial con $n = 50$ y $p = 0,3$. La Figura 5.4.1(a) muestra la distribución binomial, utilizando rayas para representar las probabilidades. Se superpone una curva normal con

$$\text{Media} = np = (50)(0,3) = 15$$

y

$$\text{Desviación típica} = \sqrt{np(1-p)} = \sqrt{(50)(0,3)(0,7)} = 3,24$$

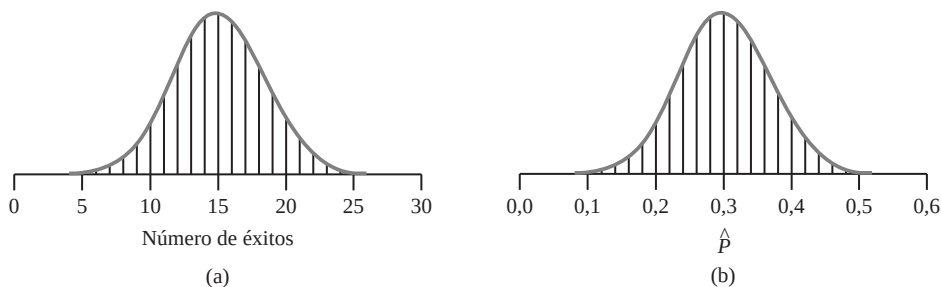


Figura 5.4.1 La aproximación normal (curva Z) a la distribución binomial (rayas negras) con $n = 50$ y $p = 0,3$

Nótese que la curva se ajusta a la distribución bastante bien. La Figura 5.4.1(b) muestra la distribución en el muestreo de $\hat{P}$ para $n = 50$ y $p = 0,3$. Se superpone una curva normal con

$$\text{Media} = p = 0,3$$

y

$$\text{Desviación típica} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0,3)(0,7)}{50}} = 0,0648$$

Nótese que la Figura 5.4.1(b) es una versión reetiquetada de la Figura 5.4.1(a).

Para ilustrar el uso de la aproximación normal, obtengamos la probabilidad de que en 50 pruebas independientes se produzcan al menos 18 éxitos. Podríamos utilizar la fórmula binomial para obtener la probabilidad de exactamente 18 éxitos en 50 pruebas y sumar después la probabilidad de exactamente 19 éxitos, exactamente 20 éxitos, y así sucesivamente:

$$\begin{aligned} \Pr\{\text{al menos 18 éxitos}\} &= {}_{50}C_{18}(0,3)^{18}(1-0,3)^{50-18} + {}_{50}C_{19}(0,3)^{19}(1-0,3)^{50-19} + \dots \\ &= 0,0772 + 0,0558 + \dots = 0,2178 \end{aligned}$$

Esta probabilidad se puede ver como el área por encima y a la derecha del «18» en la Figura 5.4.2. La aproximación normal a la probabilidad es el área correspondiente bajo la curva normal, que está sombreada en la Figura 5.4.2. El valor de z que corresponde a 18 es

$$z = \frac{18 - 15}{3,2404} = 0,93$$

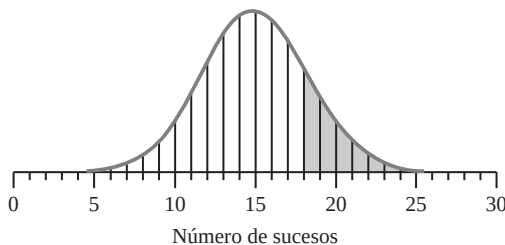


Figura 5.4.2 Aproximación normal a la probabilidad de al menos 18 éxitos

Utilizando la Tabla 3, obtenemos que el área es $1 - 0,8238 = 0,1762$, que está razonablemente cerca del valor exacto de 0,2178. Esta aproximación se puede mejorar teniendo en cuenta el hecho de que la distribución binomial es discreta y la distribución normal es continua, como veremos seguidamente. ■

La corrección de continuidad

Como hemos visto en el Capítulo 4, como la distribución normal es continua, las probabilidades se calculan como áreas bajo la curva normal, en vez de ser la altura de la curva normal en algún valor particular. Debido a esto, para calcular $\Pr\{Y = 18\}$, la probabilidad de 18 éxitos, podemos pensar que «18» cubre el espacio desde 17,5 a 18,5 y considerar así el área bajo la curva normal entre 17,5 y 18,5, como se ilustra en la Figura 5.4.3. Asimismo, para obtener una aproximación más exacta en el Ejemplo 5.4.1, podemos utilizar 17,5 en lugar de 18 al calcular el valor de z . Todo lo anterior son ejemplos de corrección de continuidad.

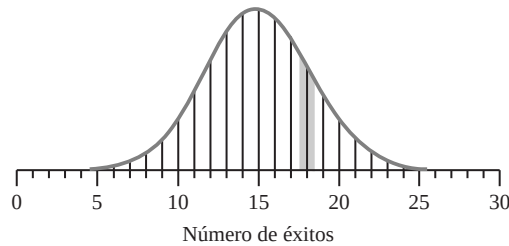


Figura 5.4.3 Aproximación normal a la probabilidad de exactamente 18 éxitos

Ejemplo 5.4.2

Aplicando la corrección de continuidad a la aproximación normal, la probabilidad de al menos 18 éxitos en 50 pruebas, cuando $p = 0,3$, se aproxima calculando

$$z = \frac{17,5 - 15}{3,2404} = 0,77$$

Utilizando la Tabla 3, obtenemos que el área por encima de 0,77 es $1 - 0,7794 = 0,2206$, que se ajusta bastante bien al valor exacto de 0,2178. Esta área se muestra en la Figura 5.4.4. ■

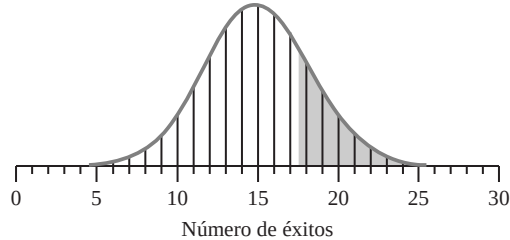


Figura 5.4.4 Aproximación normal mejorada a la probabilidad de al menos 18 éxitos

Ejemplo 5.4.3

Para ilustrar el apartado (b) del Teorema 5.4.1, supondremos de nuevo que $n = 50$ y $p = 0,3$. Consideremos el cálculo de la probabilidad de que al menos el 40 % de las 50 pruebas de un experimento binomial con $p = 0,3$ sean éxitos. Es decir, deseamos obtener $\Pr\{\hat{P} \geq 0,40\}$. La aproximación normal a esta probabilidad es el área sombreada en la Figura 5.4.5. Utilizando la corrección de continuidad, el límite del área es $\hat{p} = 19,5/50 = 0,39$, que se corresponde en la escala Z con

$$z = \frac{0,39 - 0,30}{0,0648} = 1,39$$

La aproximación resultante (utilizando la Tabla 3) es entonces

$$\Pr\{\hat{P} \geq 0,40\} \approx 1 - 0,9177 = 0,0823$$

que se ajusta muy bien al valor exacto de 0,0848 (obtenido utilizando la fórmula binomial). ■

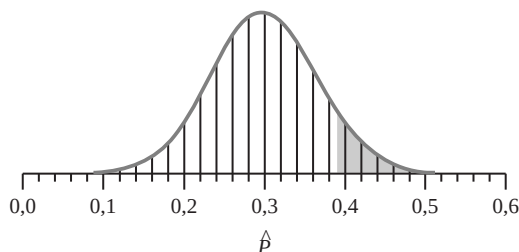


Figura 5.4.5 Aproximación normal a $\Pr\{\hat{P} \geq 0,40\}$

Observación Cualquier problema que involucre la aproximación normal a la binomial se puede resolver de dos formas: en términos de Y , utilizando el apartado (a) del Teorema 5.4.1 o en términos de $\hat{P}$ utilizando el apartado (b) de dicho teorema. Aunque es natural plantear las preguntas en términos de proporciones (por ejemplo, «¿cuánto vale $\Pr\{\hat{P} \geq 0,70\}$?»), a menudo es más fácil resolver problemas en términos de la distribución binomial Y (es decir, «¿cuánto vale $\Pr\{Y \geq 35\}$?»), particularmente cuando se utiliza la corrección de continuidad. El ejemplo siguiente ilustra el planteamiento para convertir una pregunta sobre una proporción muestral en una pregunta sobre el número de éxitos de una variable aleatoria binomial.

Ejemplo 5.4.4

Consideremos una distribución binomial con $n = 50$ y $p = 0,3$. La proporción muestral de éxitos, en las 50 pruebas, es $\hat{P}$. La Figura 5.4.1(b) muestra la distribución en el muestreo de $\hat{P}$ con una curva normal superpuesta.

Supongamos que deseamos obtener la probabilidad de que $0,24 \leq \hat{P} \leq 0,36$. Como $\hat{P} = Y/50$, la probabilidad anterior se corresponde con la probabilidad de que $0,24 \leq Y/50 \leq 0,36$, que es igual a la probabilidad de que $12 \leq Y \leq 18$. Es decir, $\Pr\{0,24 \leq \hat{P} \leq 0,36\} = \Pr\{12 \leq Y \leq 18\}$.

Sabemos que Y tiene una distribución binomial de media $= np = (50)(0,3) = 15$ y $DT = \sqrt{np(1-p)} = \sqrt{(50)(0,3)(0,7)} = 3,24$. Utilizando la corrección de continuidad obtenemos valores en la escala Z de

$$z = \frac{11,5 - 15}{3,24} = -1,08$$

y

$$z = \frac{18,5 - 15}{3,24} = 1,08$$

Entonces, utilizando la Tabla 3, tenemos que $\Pr\{0,24 \leq \hat{P} \leq 0,36\} = \Pr\{12 \leq Y \leq 18\} \approx 0,8599 - 0,1401 = 0,7198$. ■

¿Cuánto debe valer n ?

El Teorema 5.4.1 plantea que la distribución binomial se puede aproximar mediante una distribución normal si n es «grande». Es útil conocer cuánto debe valer n para que la aproximación sea adecuada. El valor de n requerido depende del valor de p . Si $p = 0,5$, entonces la distribución binomial es simétrica y la aproximación normal es muy buena incluso para valores de n tan pequeños como 10. Sin embargo, si $p = 0,1$, la distribución binomial para $n = 10$ está bastante sesgada, y se ajusta pobremente a una curva normal. Para valores de n mayores, el sesgo disminuye y la aproximación normal es mejor. Una regla práctica simple es la siguiente:

La aproximación normal a la distribución binomial es bastante buena si tanto np como $n(1-p)$ son al menos igual a 5.

Por ejemplo, si $n = 50$ y $p = 0,3$, como en el Ejemplo 5.4.4, entonces $np = 15$ y $n(1-p) = 35$, como $15 \geq 5$ y $35 \geq 5$, la regla práctica indica que la aproximación normal es bastante buena.

Ejercicios 5.4.1-5.4.13

5.4.1 Se lanza 20 veces una moneda equilibrada. Calcule la probabilidad de que en 10 de los lanzamientos salga cara y en 10 salga cruz:

- (a) Utilizando la fórmula de la distribución binomial.
- (b) Utilizando la aproximación normal con la corrección de continuidad.

5.4.2 En Estados Unidos, el 44 % de la población tiene grupo sanguíneo O. Suponga que se toma una muestra aleatoria de 12 personas. Calcule la probabilidad de que 6 de ellas tengan grupo sanguíneo O (y las otras 6 no):

- (a) Utilizando la fórmula de la distribución binomial.
- (b) Utilizando la aproximación normal.

5.4.3 En referencia al Ejercicio 5.4.2, calcule la probabilidad de que como mucho 6 de las personas tengan grupo sanguíneo O utilizando la aproximación normal:

- (a) Sin la corrección de continuidad.
- (b) Con la corrección de continuidad.

5.4.4 Una epidemióloga está planificando un estudio sobre la preponderancia del uso de anticonceptivos orales en cierta población⁹. Planea escoger una muestra aleatoria de n mujeres y utilizar la proporción muestral de las que usan anticonceptivos orales ($\hat{P}$) como estimador de la proporción poblacional (p). Suponga que de hecho $p = 0,12$. Utilice la aproximación normal (con la corrección de continuidad) para determinar la probabilidad de que ($\hat{P}$) este dentro de un intervalo de $\pm 0,03$ alrededor de p si:

- (a) $n = 100$.
- (b) $n = 200$.

[Sugerencia: si encuentra dificultades al utilizar el apartado (b) del Teorema 5.4.1, intente utilizar en su lugar el apartado (a) de dicho teorema].

5.4.5 En un estudio de cómo realiza la gente juicios de probabilidad, se planteó a unos estudiantes universitarios (sin formación en probabilidad ni estadística) la siguiente cuestión¹⁰. Una cierta ciudad está atendida por dos hospitales. En el hospital mayor nacen aproximadamente 45 bebés al día, y en el hospital más pequeño aproximadamente 15 bebés al día. Como sabe, aproximadamente el 50 % de todos los bebés son niños. El porcentaje exacto de niños, sin embargo, varía de un día a otro. Algunas veces puede ser mayor del 50 % y otras veces menor.

Durante un periodo de un año, cada hospital llevó la cuenta de los días en los que al menos el 60 % de los bebés que habían nacido eran niños. ¿Qué hospital piensa que obtuvo un número de días mayor?

- El hospital más grande.
- El hospital más pequeño.
- Aproximadamente los dos igual (es decir, dentro de un margen del 5 % de diferencia).

- (a) Imagine que es un participante en el estudio. ¿Qué respuesta escogería, basada solo en la intuición?

- (b) Determine la respuesta correcta utilizando la aproximación normal (sin la corrección de continuidad) para calcular las probabilidades apropiadas.

5.4.6 Considere el muestreo aleatorio de una población dicotómica con $p = 0,3$, y sea E el suceso de que $\hat{P}$ esta dentro de un intervalo de amplitud $\pm 0,05$ alrededor de p . Utilice la aproximación normal (sin la corrección de continuidad), para calcular la $\Pr\{E\}$ para un tamaño de muestra $n = 400$.

5.4.7 En referencia al Ejercicio 5.4.6, calcule $\Pr\{E\}$ para $n = 40$ (en vez de 400) sin la corrección de continuidad.

5.4.8 En referencia al Ejercicio 5.4.6, calcule $\Pr\{E\}$ para $n = 40$ (en vez de 400) con la corrección de continuidad.

5.4.9 Un cierto cruce entre plantas de guisantes dulce produce una progenie que tiene flores púrpura o flores blancas¹¹. La probabilidad de que una planta tenga flores púrpura es $p = \frac{9}{16}$. Suponga que se examinan n descendientes, y sea $\hat{P}$ la proporción muestral de plantas con flores púrpura. Podría suceder, por azar, que $\hat{P}$ estuviera más cerca de $\frac{1}{2}$ que de $\frac{9}{16}$. Calcule la probabilidad de que ocurra este suceso engañoso si

- (a) $n = 1$.
- (b) $n = 64$.
- (c) $n = 320$.

(Utilice la aproximación normal si la corrección de continuidad).

5.4.10 El citomegalovirus (CMV) es un virus (generalmente benigno) que infecta a la mitad de los adultos jóvenes¹². Si se toma una muestra aleatoria de 10 adultos jóvenes, calcule la probabilidad de que entre el 30 % y el 40 % (ambos inclusive) de los adultos jóvenes de la muestra tengan CMV:

- (a) Utilizando la fórmula de la distribución binomial.
- (b) Utilizando la aproximación normal con la corrección de continuidad.

5.4.11 En una cierta población de mejillones (*Mytilus edulis*), el 80 % de los individuos están infectados con un parásito intestinal¹³. Un biólogo marino planea examinar 100 mejillones escogidos aleatoriamente de la población. Calcule la probabilidad de que el 85 % o más de los mejillones de la muestra estén infectados, utilizando la aproximación normal sin la corrección de continuidad.

5.4.12 En referencia al Ejercicio 5.4.11, calcule la probabilidad de que el 85 % o más de los mejillones muestreados estén infectados, utilizando la aproximación normal con la corrección de continuidad.

5.4.13 En referencia al Ejercicio 5.4.11, suponga que el biólogo toma una muestra aleatoria de tamaño 50. Calcule la probabilidad de que menos de 35 de los mejillones de la muestra estén infectados, utilizando la aproximación normal:

- (a) Sin la corrección de continuidad.
- (b) Con la corrección de continuidad.

5.5 Perspectiva

En este capítulo hemos presentado el concepto de distribución muestral y nos hemos centrado en la distribución muestral de $\bar{Y}$. Por supuesto, existen otras muchas distribuciones en el muestrales importantes, como la distribución muestral de la desviación típica y la distribución muestral de la mediana.

Echemos otro vistazo al modelo de muestreo aleatorio a la luz del Capítulo 5. Como hemos visto, una muestra *aleatoria* no es necesariamente una muestra *representativa**. Pero utilizando las distribuciones muestrales, se puede especificar el grado de representatividad que se puede esperar de una muestra aleatoria. Por ejemplo, es intuitivamente convincente que una muestra aleatoria más grande es más probable que sea más representativa que una muestra aleatoria más pequeña de la misma población. En las Secciones 5.1 y 5.2 vimos cómo el concepto de distribución muestral puede hacer esta vaga intuición más precisa, especificando la probabilidad de que una muestra aleatoria alcance un grado de representatividad concreto. Por tanto, las distribuciones muestrales proporcionan lo que se ha denominado «certeza sobre la incertidumbre»¹⁴.

En el Capítulo 6 veremos por primera vez cómo la teoría de las distribuciones muestrales se puede aplicar al uso práctico en el análisis de datos. Encontraremos que, aunque los cálculos del Capítulo 5 parecen requerir el conocimiento de cantidades desconocidas (como μ o σ), cuando se analizan datos se puede no obstante estimar la magnitud probable del error debido al muestreo utilizando solo la información contenida en la propia muestra.

Además de su aplicación en el análisis de datos, las distribuciones muestrales proporcionan una base para comparar los méritos relativos de diferentes métodos de análisis. Por ejemplo, consideremos el muestreo de una población normal de media μ . Por supuesto, la media muestral $\bar{Y}$ es un estimador de μ . Pero como una distribución normal es simétrica, μ es también la mediana poblacional, de forma que la *mediana* muestral es también un estimador de μ . ¿Cómo podemos entonces decidir qué estimador es mejor? Esta pregunta se puede responder en términos de distribuciones muestrales, como sigue: los estadísticos han determinado que, si la población es normal, la mediana muestral es inferior a la media muestral en el sentido de que su distribución muestral tiene una desviación típica mayor que $\frac{\sigma}{\sqrt{n}}$.

En consecuencia, la mediana muestral es menos eficiente (como estimador de μ) que la media muestral. Dado un determinado tamaño de la muestra n , la mediana muestral proporciona menos información sobre μ que la media muestral. (Si la población no es normal, sin embargo, la mediana muestral puede ser mucho más eficiente que la media).

* Sin embargo, es cierto que algunas veces el investigador puede forzar que la muestra sea representativa con respecto a una variable (no la que está bajo estudio) cuya distribución poblacional sea conocida. Por ejemplo, utilizando un muestreo aleatorio por estratos como se presenta en la Sección 1.3. Los métodos de análisis dados en este libro, sin embargo, solo son apropiados para muestras aleatorias simples y no se pueden aplicar sin las modificaciones adecuadas.

Ejercicios suplementarios 5.S.1-5.S.12

(Nota: los ejercicios precedidos por un asterisco se refieren a las secciones opcionales).

5.S.1 En un experimento agrícola, un campo grande de trigo se dividió en muchas parcelas (siendo cada parcela de 7×100 pies) y se midió para cada parcela el rendimiento de grano. Dichos rendimientos por parcela siguieron una distribución aproximadamente normal de media 88 lb y desviación típica 7 lb (como en el Ejercicio 4.3.5). Sea $\bar{Y}$ el rendimiento medio de cinco parcelas escogidas aleatoriamente en el campo. Calcule $\Pr\{\bar{Y} > 90\}$.

5.S.2 Considere tomar una muestra aleatoria de tamaño 14 de una población de estudiantes de una cierta universidad y medir la presión sanguínea diastólica de cada uno de ellos. En este contexto, explique lo que significa la distribución muestral de la media.

5.S.3 En relación con el Ejercicio 5.S.2, suponga que la media poblacional es de 70 mmHg y que la desviación típica poblacional es de 10 mmHg. Si el tamaño de la muestra es de 14, ¿cuál es la desviación típica de la distribución muestral de la media?

5.S.4 Las alturas de los hombres en una cierta población siguen una distribución normal de media 69,7 pulgadas y desviación típica 2,8 pulgadas¹⁵.

- Si se escoge aleatoriamente un hombre de la población, calcule la probabilidad de que su altura sea mayor que 72 pulgadas.
- Si se escogen aleatoriamente dos hombres de la población, calcule la probabilidad de que (i) ambos midan más de 72 pulgadas; (ii) su altura media sea mayor que 72 pulgadas.

5.5.5 Suponga que una botánica cultiva muchas berenjenas en parcelas individuales, todas ellas tratadas idénticamente y dispuestas en grupos de cuatro parcelas en los bancos del invernadero. Tras 30 días de crecimiento, mide el área total Y de las hojas de cada planta. Suponga que la distribución poblacional de Y es aproximadamente normal de media $= 800 \text{ cm}^2$ y $DT = 900 \text{ cm}^2$ ¹⁶.

- ¿Qué porcentaje de las plantas en la población tendrán áreas de las hojas entre 750 cm^2 y 850 cm^2 ?
- Suponga que cada grupo de cuatro plantas se puede considerar como una muestra aleatoria de la población. ¿Qué porcentaje de los grupos tendrá un área media de hoja de grupo entre 750 cm^2 y 850 cm^2 ?

5.5.6 En relación con el Ejercicio 5.5.5, en un invernadero real, ¿qué factores podrían tender a invalidar el supuesto de que cada grupo de plantas se puede considerar como una muestra aleatoria de la misma población?

***5.5.7** Considere que se toma una muestra aleatoria de tamaño 25 de una población en la que el 42 % de las personas tiene grupo sanguíneo A. ¿Cuál es la probabilidad de que la proporción muestral de personas con grupo sanguíneo A sea mayor que 0,44? Utilice la aproximación normal a la binomial con la corrección de continuidad.

5.5.8 La actividad de una cierta enzima se mide contando las emisiones de una molécula etiquetada radiactivamente. En un cierto espécimen de tejido, el recuento de emisiones en periodos de tiempo consecutivos de 10 segundos se pueden considerar (aproximadamente) como observaciones independientes repetidas de una distribución normal (como en el Ejercicio 4.5.1). Suponga que el valor medio del recuento de 10 segundos de un cierto espécimen de tejido es 1.200 y la desviación típica es 35. Para ese espécimen, sea Y la variable que representa el recuento en 10 segundos e $\bar{Y}$ la media de seis recuentos de 10 segundos. Tanto Y como $\bar{Y}$ son insesgadas (ambas tienen una media de 1.200), pero esto no implica que sean igualmente buenas. Calcule $\Pr\{1.175 \leq Y \leq 1.225\}$ y $\Pr\{1.175 \leq \bar{Y} \leq 1.225\}$, y compárelas. ¿Indica la compa-

ración que contar durante un minuto y dividir por 6 tendería a dar un resultado más preciso que simplemente contar durante un único periodo de 10 segundos? ¿Cómo?

5.5.9 En una cierta población de laboratorio de ratones, los pesos a los 20 días de edad siguen aproximadamente una distribución normal con peso medio igual a 8,3 g y desviación típica igual a 1,7 g¹⁷. Suponga que se pesan muchas camadas de 10 ratones cada una. Si cada camada se puede considerar como una muestra aleatoria de la población, ¿qué porcentaje de las camadas tendrán un peso total de 90 g o más? (*Sugerencia*: ¿cómo se relaciona el peso total de una camada con el peso medio de sus miembros?).

5.5.10 En referencia al Ejercicio 5.5.9, en realidad, ¿qué factores tenderían a invalidar el supuesto de que cada camada se puede considerar como una muestra aleatoria de la misma población?

5.5.11 Considere que se toma una muestra aleatoria de tamaño 25 de una población de plantas, se mide el peso de cada planta y se suman los pesos para obtener un total muestral. En este contexto, explique qué quiere decir la distribución muestral del total.

5.5.12 La anchura del cráneo de una cierta población de roedores sigue una distribución normal con desviación típica de 10 mm. Sea $\bar{Y}$ la media de la anchura craneal de una muestra aleatoria de 64 individuos de esta población, y sea μ la media poblacional de dicha anchura craneal.

- Suponga que $\mu = 50$ mm. Calcule la probabilidad de que $\bar{Y}$ esté dentro de un intervalo de amplitud ± 2 mm alrededor de μ .
- Suponga que $\mu = 100$ mm. Calcule la probabilidad de que $\bar{Y}$ esté dentro de un intervalo de amplitud ± 2 mm alrededor de μ .
- Suponga que μ es desconocida. ¿Se puede calcular la probabilidad de que $\bar{Y}$ esté dentro de un intervalo de amplitud ± 2 mm alrededor de μ ? Si es así, calcúlela. Si no, explique por qué no.

Notas

- Datos del 2003-2004 National Health and Nutrition Examination Survey, que se pueden encontrar en www.denofinquiry.com/nhanes/source/choose.php
- Población ficticia pero realista. Véase el Ejemplo 2.2.11.
- La media y la desviación típica son realistas para mujeres americanas de edades entre 25 y 34 años. Véase O'Brien, R. J. y Drizd, T. A. (1981). Basic data on spirometry in adults 25-74 years of age: United States, 1971-75. *U.S. National Center for Health Statistics, Vital and Health Statistics*, Serie 11, n.º 222. Washington, D.C.: U.S. Department of Health and Human Services. El supuesto de normalidad puede ser o no ser realista.

- Adaptado de datos proporcionados en Sebens, K. P. (1981). Recruitment in a sea anemone population: juvenile substrate becomes adult prey. *Science* **213**, 785-787.
- Datos ficticios pero realistas. Adaptado de la distribución para hombres entre 45 y 59 años de edad dada en Roberts, J. (1975). Blood pressure of persons 18-74 years, United States, 1971-72. *U.S. National Center for Health Statistics, Vital and Health Statistics*, Serie 11, n.º 150. Washington, D.C.: U.S. Department of Health, Education and Welfare.
- Basado en datos de Roberts, J. D., et al. (1997). Inhaled nitric oxide and persistent pulmonary hypertension of the newborn. *New England Journal of Medicine* **336**, 605-610.

7. La distribución de la Figura 5.3.1 se basa en datos dados en Zeleny, C. (1922). The effect of selection for eye facet number in the white bar-eye race of *Drosophila melanogaster*. *Genetics* 7, 1-115. Los datos se representan en Falconer, D. S. (1981). *Introduction to Quantitative Genetics*, 2.^a ed. New York: Longman, p. 97.
8. La distribución de la Figura 5.3.3 está adaptada de los datos descritos por Bradley, J. V. (1980). Nonrobustness in one-sample Z and t tests: A large-scale sampling study. *Bulletin of the Psychonomic Society* 15 (1), 29-32, utilizado con permiso de la Psychonomic Society, Inc.; y Bradley, J. V. (1977). A common situation conducive to bizarre distribution shapes. *American Statistician* 31, 147-150. La distribución de Bradley incluía picos adicionales, ya que algunas veces el sujeto accionaba el botón más de una vez en una sola prueba.
9. Situación ficticia pero realista, adaptada de datos dados en Bradley, D. D., Krauss, R. M., Petite, D. B., Ramcharin, S. y Wingird, I. (1978). Serum high-density lipoprotein cholesterol in women using oral contraceptives, estrogens, and progestins. *New England Journal of Medicine* 299, 17-20.
10. Kahneman, D. y Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3, 430-454.
11. Strickberger, M. W. (1976). *Genetics*, 2.^a ed. New York: Macmillan, p. 206.
12. www.cureresearch.com/artic/other_important_std_niaid_fact_sheet_niaid.htm
13. Situación ficticia pero realista. Véase Waugh, G. D. (1954). The occurrence of *Mytilicola intestinalis* (Steuer) on the east coast of England. *Journal of Animal Ecology* 23, 364-367.
14. Mosteller, F. y Tukey, J. W. (1977). *Data Analysis and Regression*. Reading, Mass.: Addison-Wesley, p. 25.
15. Población ficticia pero realista, parecida a la población de jóvenes americanos entre 18 y 24 años de edad, como se describe en Abraham, S., Johnson, C. L. y Najjar, M. F. (1979). Weight and height of adults 18-74 years of age: United States 1971-1974. *U.S. National Center for Health Statistics*, Serie 11, n.º 211. Washington, D.C.: U.S. Department of Health, Education and Welfare.
16. La media y la desviación típica son realistas, basadas en datos no publicados proporcionados por cortesía de J. Y. Ustimer y C. A. Mitchell. El supuesto de normalidad puede ser o no ser realista.
17. La media y la desviación típica son realistas, basadas en datos no publicados proporcionados por cortesía de S. Newman y D. L. Harris. El supuesto de normalidad puede ser o no ser realista.

INTERVALOS DE CONFIANZA

Objetivos

En este capítulo comenzaremos un estudio formal de la inferencia estadística:

- introduciremos el concepto de error típico para cuantificar el grado de incertidumbre de una cantidad estimada y lo compararemos con la desviación típica;
- demostraremos la construcción e interpretación de intervalos de confianza para medias;
- proporcionaremos un método para determinar qué tamaño de la muestra es necesario para conseguir un nivel deseado de exactitud;
- consideraremos las condiciones bajo las que el uso del intervalo de confianza es válido;
- presentaremos el error típico de una diferencia de medias muestrales;
- demostraremos la construcción e interpretación de intervalos de confianza para diferencias de medias.

6.1 Estimación estadística

En este capítulo emprenderemos nuestra primera aventura sustancial en el mundo de la inferencia estadística. Recuerdese que la inferencia estadística se basa en el modelo de muestreo aleatorio: vemos nuestros datos como una muestra aleatoria de alguna población, y utilizamos la información de la muestra para inferir hechos sobre la población. La estimación estadística es una forma de inferencia estadística en la que utilizamos los datos para (1) determinar un estimador de alguna característica de la población y (2) evaluar la precisión del estimador. Consideremos un ejemplo.

Ejemplo 6.1.1 Alas de mariposas

Como parte de un estudio más amplio de la composición del cuerpo, unos investigadores capturaron 14 mariposas Monarch macho en el Océano Dunes State Park de California y midieron el área de las alas (en cm^2). Los datos se muestran en la Tabla 6.1.1¹.

Tabla 6.1.1 Área de alas de mariposas Monarch macho				
Área de las alas (cm^2)				
33,9	33,0	30,6	36,6	36,5
34,0	36,1	32,0	28,0	32,0
32,2	32,2	32,3	30,0	

Para estos datos, la media y la desviación típica son

$$\bar{y} = 32,8143 \approx 32,81 \text{ cm}^2 \quad \text{y} \quad s = 2,4757 \approx 2,48 \text{ cm}^2$$

Supongamos que consideramos las 14 observaciones como una muestra aleatoria de una población. La población se puede describir (entre otras cosas) mediante su media, μ , y su desviación típica, σ . Podríamos definir verbalmente μ y σ de la siguiente forma:

μ = La media (poblacional) del área de las alas de mariposas Monarch macho en la región de Océano Dunes.

σ = La DT (poblacional) del área de las alas de mariposas Monarch macho en la región de Océano Dunes.

Es natural estimar μ mediante la media muestral y σ mediante la desviación típica muestral. Por tanto, para los datos de las 14 mariposas,

32,81 es una estimación de μ .

2,48 es una estimación de σ .

Sabemos que estas estimaciones están sujetas a error de muestreo. Nótese que no estamos hablando meramente de error de medida. No importa la exactitud con la que se haya medido cada ala de mariposa. La información de la muestra es imperfecta debido al hecho de que se han medido solo 14 alas, en vez de medir toda la población de mariposas. ■

En general, dada una muestra de observaciones de una variable cuantitativa Y , la media y la DT muestrales son estimadores de la media y DT poblacionales:

$\bar{y}$ es un estimador de μ .

s es un estimador de σ .

La notación para estas medias y DT se resume esquemáticamente en la Figura 6.1.1.

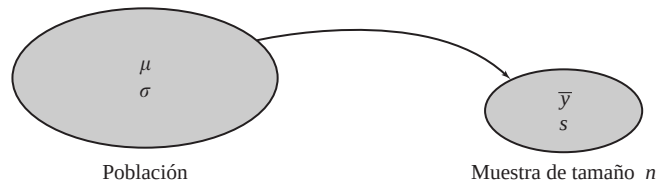


Figura 6.1.1 Notación de medias y DT de la muestra y la población

Nuestro objetivo es estimar μ . Veremos cómo evaluar la fiabilidad o la precisión de este estimador, y cómo planificar un estudio lo suficientemente grande para obtener una precisión determinada.

6.2 Error típico de la media

Es intuitivamente razonable pensar que la media muestral $\bar{y}$ sea un estimador de μ . No es tan obvio cómo determinar la fiabilidad del estimador. Como estimador de μ , la media muestral $\bar{y}$ es imprecisa en el sentido de que está afectada por el error de muestreo. En la Sección 5.3 vimos que el módulo del error de muestreo (es decir, la cantidad de discrepancia entre $\bar{y}$ y μ), está descrita (en un sentido probabilístico) por la distribución muestral de $\bar{Y}$. La desviación típica de la distribución muestral de $\bar{Y}$ es

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

Como s es una estimación de σ , una estimación natural de $\frac{\sigma}{\sqrt{n}}$ sería $\frac{s}{\sqrt{n}}$. Esta cantidad se denomina **error típico de la media**. Lo denominaremos $ET_{\bar{y}}$, o algunas veces, simplemente ET^* .

* Algunos estadísticos prefieren utilizar el término «error típico» para $\sigma/\sqrt{n}$ y denominar a $s/\sqrt{n}$ «error típico estimado».

Definición El **error típico de la media** se define como

$$ET_{\bar{y}} = \frac{s}{\sqrt{n}}$$

El siguiente ejemplo ilustra esta definición.

Ejemplo 6.2.1 Alas de mariposas

Para los datos de las mariposas Monarch del Ejemplo 6.1.1, tenemos que $n = 14$, $\bar{y} = 32,8143 \approx 32,81 \text{ cm}^2$ y $s = 2,4757 \approx 2,48 \text{ cm}^2$. El error típico de la media es

$$\begin{aligned} ET_{\bar{y}} &= \frac{s}{\sqrt{n}} \\ &= \frac{2,4757}{\sqrt{14}} = 0,6617 \text{ cm}^2, \text{ que redondearemos a } 0,66 \text{ cm}^2* \end{aligned}$$

Como hemos visto, el ET es un estimador de $\sigma_{\bar{y}}$. A un nivel más práctico, el ET se puede interpretar en términos del error de muestreo esperado. Hablando en términos generales, la diferencia entre $\bar{y}$ y μ raramente es mayor que unos pocos errores típicos. De hecho, podemos esperar que $\bar{y}$ esté muy a menudo dentro de un intervalo de un error típico alrededor de μ . Cuanto menor sea el ET, más precisa será la estimación. Nótese sin embargo que el ET incorpora los dos factores que afectan a la fiabilidad: (1) la variabilidad inherente de las observaciones (que se expresa con s) y (2) el tamaño de la muestra (n).

Error típico frente a desviación típica

Los términos «error típico» y «desviación típica» se confunden a veces. Es extremadamente importante distinguir entre el error típico (ET) y la desviación típica (s , o DT). Estas dos cantidades describen aspectos completamente diferentes de los datos. La DT describe la dispersión de los datos, mientras que el ET describe la falta de fiabilidad (debida al error de muestreo) de la *media* de la muestra como un estimador de la *media* de la población. Consideremos un ejemplo concreto.

Ejemplo 6.2.2 Pesos de corderos al nacer

Una genetista pesó 28 corderos hembra al nacer. Todos los corderos nacieron en abril, todos eran de la misma raza (Rambouillet) y todos fueron nacimientos de un solo cordero (no había gemelos). La dieta y otras condiciones experimentales fueron las mismas para todos los progenitores. Los pesos al nacer se muestran en la Tabla 6.2.1².

Tabla 6.2.1 Pesos al nacer de 28 corderos Rambouillet						
Peso al nacer (kg)						
4,3	5,2	6,2	6,7	5,3	4,9	4,7
5,5	5,3	4,0	4,9	5,2	4,9	5,3
5,4	5,5	3,6	5,8	5,6	5,0	5,2
5,8	6,1	4,9	4,5	4,8	5,4	4,7

* Resumen de la estadística del redondeo.

Para dar el resultado de la media, desviación típica y error típico de la media, se recomienda el siguiente procedimiento:

1. Redondear el ET a dos dígitos significativos.
2. Redondear $\bar{y}$ y s para ajustar el ET con respecto a la posición decimal del último dígito significativo. (El concepto de dígito significativo se considera en el Apéndice 6.1). Por ejemplo, si el ET se redondea a la centésima más cercana, entonces $\bar{y}$ y s se deben redondear también a la centésima más cercana.

La media de esos datos es $\bar{y} = 5,17$ kg, la desviación típica es $s = 0,65$ kg y el error típico es $ET = 0,12$ kg. La DT, s , describe la variabilidad de los pesos al nacer de todos los corderos de la muestra, mientras que el ET indica la variabilidad asociada con la media muestral (5,17 kg), vista como un estimador de la media poblacional de los pesos al nacer. Esta distinción se recalca en la Figura 6.2.1, que muestra un histograma de los datos de pesos de corderos al nacer. La DT se indica como una desviación alrededor de $\bar{y}$, mientras que el ET se indica como una variación asociada a la propia $\bar{y}$. ■

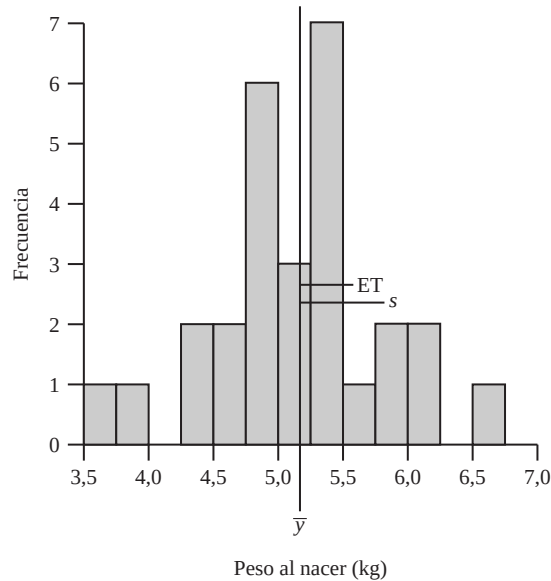


Figura 6.2.1 Pesos al nacer de 28 corderos

Otra forma de recalcar la diferencia entre ET y DT es considerar muestras de varios tamaños. A medida que el tamaño de la muestra crece, la media muestral y la DT tienden a aproximarse cada vez más a la media y DT de la población. De hecho, la distribución de los datos tiende a aproximarse a la distribución poblacional. Sin embargo, el error típico tiende a disminuir cuando n crece. Cuando n es muy grande, el ET es muy pequeño, de forma que la media muestral es un estimador muy preciso de la media poblacional. El ejemplo siguiente ilustra este efecto.

Ejemplo 6.2.3 Pesos de corderos al nacer

Supongamos que consideramos los datos de pesos al nacer del Ejemplo 6.2.2, como una muestra de tamaño $n = 28$ de una población, y que consideramos lo que sucedería si extrajéramos muestras más grandes de la misma

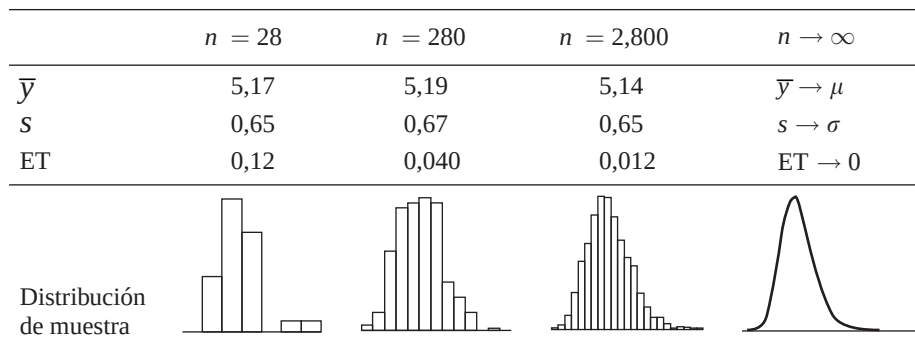


Figura 6.2.2 Muestras y tamaños en la población de corderos

población: es decir, si midiéramos los pesos al nacer de corderos hembra Rambouillet adicionales bajo las mismas condiciones. La Figura 6.2.2 muestra el tipo de resultados que podríamos esperar. Los valores que se presentan son ficticios pero realistas. Para n muy grande, $\bar{y}$ y s están muy cerca de μ y σ , siendo

μ = Peso medio al nacer de corderos Rambouillet hembra nacidos bajo las condiciones descritas.

σ = Desviación típica de los pesos al nacer de corderos Rambouillet hembra nacidos bajo las condiciones descritas. ■

Representación gráfica de ET y DT

La claridad y el impacto de un informe científico se pueden mejorar enormemente mediante gráficos de datos que estén bien diseñados. Los datos se pueden representar gráficamente o en una tabla. Presentaremos brevemente algunas de las opciones.

Consideremos en primer lugar la representación gráfica de los datos. He aquí un ejemplo.

Ejemplo 6.2.4 MAO y esquizofrenia

La enzima monoamina oxidasa (MAO) es de interés en el estudio del comportamiento humano. Las Figuras 6.2.3 y 6.2.4 muestran medidas de actividad MAO en las plaquetas de la sangre de cinco grupos de personas: los grupos I, II y III son tres categorías diagnósticas de pacientes esquizofrénicos (véase el Ejemplo 1.1.4), y los grupos IV y V son controles de hombres y mujeres sanos³. Los valores de actividad MAO se expresan en nmol de producto de benzilaldehído por 10^8 plaquetas por hora. En las dos Figuras 6.2.3 y 6.2.4 los puntos (a) o las barras (b) representan las medias de los grupos. Las líneas verticales representan $\pm ET$ en la Figura 6.2.3 y $\pm DT$ en la Figura 6.2.4.

Las Figuras 6.2.3 y 6.2.4 expresan información muy diferente. La Figura 6.2.1 expresa (1) el valor medio de MAO de cada grupo y (2) la fiabilidad de la media de cada grupo, vista como un estimador de su respectiva media poblacional. La Figura 6.2.4 expresa (1) el valor medio de MAO en cada grupo y (2) la variabilidad de la MAO dentro de cada grupo. Por ejemplo, el grupo V muestra mayor variabilidad de la MAO que el grupo I (Figura 6.2.4), pero tiene un error típico mucho menor (Figura 6.2.3) debido a que es un grupo mucho mayor.

La Figura 6.2.3 invita a quien la observa a comparar las medias y da alguna indicación de la fiabilidad de las comparaciones (pero una exposición completa de las comparaciones de dos o más medias deberá esperar hasta el Capítulo 7 y capítulos posteriores). La Figura 6.2.4 invita a quien la observa a comparar las medias y a comparar también las desviaciones típicas. Es más, la Figura 6.2.4 proporciona a quien la observa alguna información sobre la extensión del solapamiento de los valores de la MAO en los diversos grupos. Por ejemplo, consideremos los grupos IV y V. Aunque parecen estar bastante «separados» en la Figura 6.2.3, se puede ver fácilmente en la Figura 6.2.4 que hay un solapamiento considerable en los valores individuales de la MAO en los dos grupos. ■

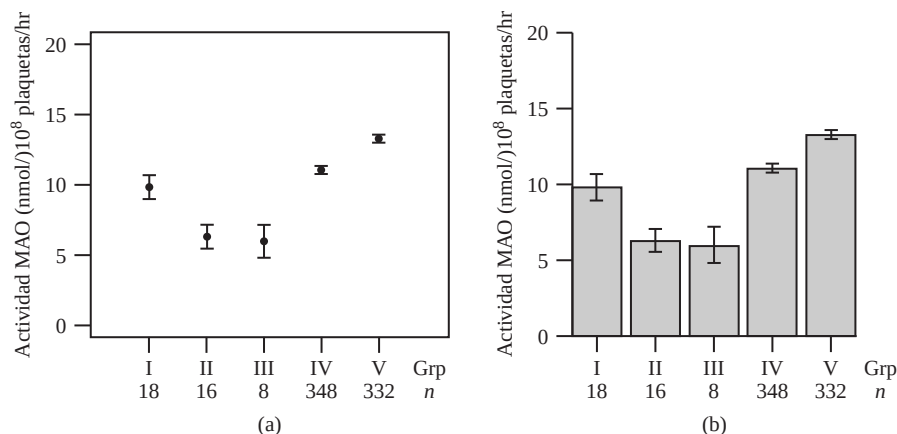


Figura 6.2.3 Datos de la MAO presentados como $\bar{y} \pm ET$ utilizando (a) una gráfica de intervalos y (b) un diagrama de barras con barras de error típico

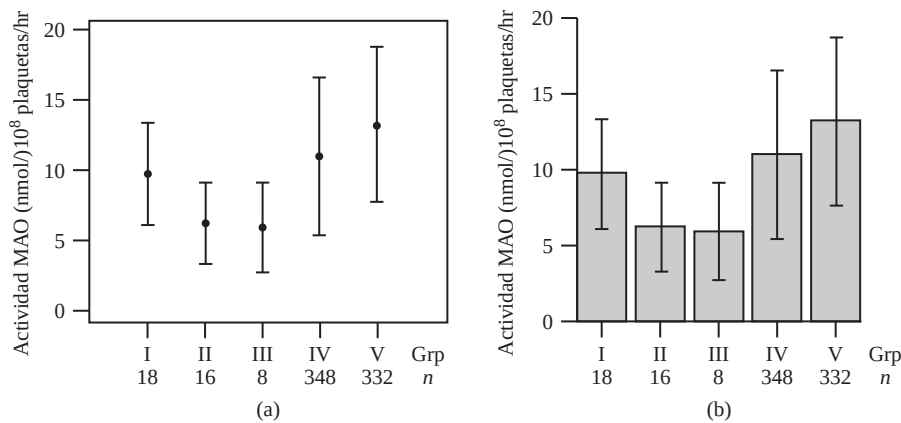


Figura 6.2.4 Datos de la MAO presentados como $\bar{y} \pm ET$ utilizando (a) una gráfica de intervalos y (b) un diagrama de barras con barras de desviación típica

Aunque hemos presentado los datos de MAO utilizando cuatro gráficas separadas en las Figuras 6.2.3 y 6.2.4, en general escogeríamos solamente una de ellas para publicarla en un informe. Escoger entre la gráfica de intervalos o el diagrama de barras es una cuestión de preferencias y estilo personales. Y, como se mencionó previamente, escoger si las barras de intervalos representan la DT o el ET dependerá si deseamos recalcar una comparación de las medias (ET) o, de forma más simple, un resumen de la variabilidad de nuestros datos observados (DT)*.

En algunos informes científicos, los datos se resumen en tablas, en vez de gráficamente. La Tabla 6.2.2 muestra un resumen en forma de tabla de los datos de MAO del Ejemplo 6.2.4. Como con las gráficas anteriores, al presentar formalmente los resultados, generalmente se muestra o bien la DT o el ET, pero no ambos.

Tabla 6.2.2 Actividad MAO en cinco grupos de personas				
Actividad MAO (nmol/10 ⁸ plaquetas/hr)				
Grupo	n	Media	ET	DT
I	18	9,81	0,85	3,62
II	16	6,28	0,72	2,88
III	8	5,97	1,13	3,19
IV	348	11,04	0,30	5,59
V	332	13,29	0,30	5,50

Ejercicios 6.2.1-6.2.7

6.2.1 Un farmacólogo midió la concentración de dopamina en los cerebros de varias ratas. La concentración media fue de 1,269 ng/g, y la desviación típica fue de 145 ng/g⁴. Indique cuál fue el error típico de la media si:

- (a) Se midieron 8 ratas.
- (b) Se midieron 30 ratas.

6.2.2 Un agrónomo midió las alturas de n plantas de maíz⁵. La altura media fue de 220 cm y la desviación típica de 15 cm. Calcule el error típico de la media si

- (a) $n = 25$
- (b) $n = 100$

6.2.3 Al evaluar una cosecha de forraje, es importante medir la concentración de varios constituyentes del tejido de la

* Para presentar una gráfica ligeramente más simple, en los diagramas de barras a menudo solo se representa la parte superior de las barras de error (ET o DT).

planta. En un estudio de la fiabilidad de estas medidas, se secó un montón de alfalfa, se trituró y se pasó por un fino tamiz. Se analizó el contenido de residuo insoluble en cinco pequeñas partes iguales (de 0,3 g) de dicha alfalfa⁶. Los resultados (g/kg) fueron los siguientes:

10,0 8,9 9,1 11,7 7,9

Para estos datos, calcule la media, la desviación típica y el error típico de la media.

6.2.4 Un zoólogo midió la longitud de la cola en 86 individuos, todos en el grupo de un año de edad, del ratón ciervo *Peromyscus*. La longitud media fue de 60,43 mm y la desviación típica fue de 3,06 mm. La tabla presenta una distribución de frecuencias de los datos⁷.

Longitud de la cola (mm)	N.º de ratones
[52, 54)	1
[54, 56)	3
[56, 58)	11
[58, 60)	18
[60, 62)	21
[62, 64)	20
[64, 66)	9
[66, 68)	2
[68, 70)	1
Total	86

(a) Calcule el error típico de la media.

(b) Construya un histograma de los datos e indique en dicho histograma los intervalos $\bar{y} \pm DT$ e $\bar{y} \pm ET$ (véase la Figura 6.2.1).

6.2.5 En referencia a los datos de ratones del Ejercicio 6.2.4, suponga que el zoólogo fuera a medir 500 animales adicionales de la misma población. Basándose en los datos del Ejercicio 6.2.4:

(a) ¿Cómo podría predecir que sería la desviación típica de las 500 nuevas medidas?

(b) ¿Cómo podría predecir que sería el error típico de la media de las 500 nuevas medidas?

6.2.6 En un informe de un estudio farmacológico, los animales experimentales se describieron como sigue⁸: «se inyectó a ratas cuyo peso era de 150 ± 10 g...» con un cierto compuesto químico, y después se realizaron ciertas medidas en la ratas. Si el autor intenta expresar el grado de homogeneidad del grupo de animales experimentales, entonces los 10 g, ¿deberían ser la DT o el ET? Explique su respuesta.

6.2.7 En cada uno de los siguientes casos, decida si la descripción se ajusta mejor a la DT o al ET.

(a) Esta cantidad es una medida de la precisión de la media muestral como estimador de la media poblacional.

(b) Esta cantidad tiende a permanecer igual cuando el tamaño de la muestra aumenta.

(c) Esta cantidad tiende a disminuir cuando el tamaño de la muestra aumenta.

6.3 Intervalo de confianza para μ

En la Sección 6.2 dijimos que el error típico de la media (el ET) mide lo lejos que es probable que esté $\bar{y}$ de la media poblacional μ . En esta sección precisaremos esa idea.

Intervalo de confianza para μ : idea básica

La Figura 6.3.1 es un dibujo de un hombre invisible paseando a su perro. El perro, que es visible, tiene una correa invisible con resorte. La tensión sobre el resorte es tal que el perro está dentro de una distancia de 1 ET del hombre aproximadamente las dos terceras partes del tiempo. El perro está dentro de una distancia de dos errores típicos del hombre el 95 % del tiempo. Sólo el 5 % del tiempo el perro está a una distancia mayor de 2 ET del hombre (a menos que la correa se rompa, en cuyo caso el perro podría estar en cualquier parte). Podemos ver al perro, pero nos gustaría saber dónde está el hombre. Como el hombre y el perro están generalmente a una distancia menor de 2 ET entre sí, podemos tomar el intervalo «perro $\pm 2 \times ET$ » como un intervalo que típicamente incluiría al hombre. De hecho, podría decirse que tenemos una confianza del 95 % de que el hombre esté en este intervalo.



Figura 6.3.1 Hombre invisible paseando a su perro

Esta es la idea básica del intervalo de confianza. Nos gustaría saber el valor de la media poblacional μ (que corresponde al hombre) pero no podemos verla directamente. Lo que *podemos* ver es la media muestral $\bar{y}$ (que corresponde

al perro). Utilizamos lo que podemos ver, $\bar{y}$, junto con el error típico, que podemos calcular a partir de los datos, como una forma de construir un intervalo que esperamos incluya lo que no podemos ver, la media poblacional μ . Llamamos intervalo de confianza al 95 % a ese intervalo «posición del perro $\pm 2 \times \text{ET}$ ». [Todo esto depende de tener un modelo que sea correcto: hemos dicho que si el resorte se rompe, entonces saber dónde está el perro no nos dice mucho sobre dónde está el hombre. Asimismo, si nuestro modelo estadístico es incorrecto (por ejemplo, si tenemos una muestra sesgada), entonces ¡conocer $\bar{y}$ no nos dirá mucho sobre μ !].

Intervalo de confianza para μ : matemáticas

En la analogía del hombre invisible*, dijimos que el perro está dentro de una distancia de 1 ET del hombre aproximadamente las dos terceras partes del tiempo y dentro de una distancia de 2 ET del hombre el 95 % del tiempo. Esto se basa en la idea de la distribución muestral de $\bar{Y}$ cuando tenemos una muestra aleatoria de una distribución normal. Si Z es la variable aleatoria normal tipificada, entonces la probabilidad de que Z este entre ± 2 es aproximadamente el 95 %. De forma más precisa, $\Pr\{-1,96 < Z < 1,96\} = 0,95$. En el Capítulo 5 vimos que si Y sigue una distribución normal, entonces

$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$ tiene una distribución normal tipificada (Z), de forma que

$$\Pr\left\{-1,96 < \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < 1,96\right\} = 0,95 \quad (6.3.1)$$

Por tanto,

$$\Pr\{-1,96 \times \sigma/\sqrt{n} < \bar{Y} - \mu < 1,96 \times \sigma/\sqrt{n}\} = 0,95$$

y

$$\Pr\{-\bar{Y} - 1,96 \times \sigma/\sqrt{n} < -\mu < -\bar{Y} + 1,96 \times \sigma/\sqrt{n}\} = 0,95$$

por lo que

$$\Pr\{\bar{Y} - 1,96 \times \sigma/\sqrt{n} < \mu < \bar{Y} + 1,96 \times \sigma/\sqrt{n}\} = 0,95$$

Es decir, el intervalo

$$\bar{Y} \pm 1,96 \frac{\sigma}{\sqrt{n}} \quad (6.3.2)$$

contendrá a μ en el 95 % de las muestras.

El intervalo (6.3.2) no se puede utilizar para el análisis de datos porque contiene una cantidad (concretamente σ) que no se puede determinar a partir de los datos. Si sustituimos σ por su estimador (concretamente s), entonces podemos calcular un intervalo a partir de los datos, pero entonces ¿qué sucede con la interpretación del 95 %? Afortunadamente, sucede que hay una vía de escape de este dilema. La vía de escape fue descubierta por un científico británico de nombre W. S. Gosset, que era empleado de la Fábrica de Cerveza Guinness. Publicó sus descubrimientos en 1908 bajo el seudónimo «Student», y el método ha llevado ese nombre desde entonces⁹. «Student» descubrió que, si los datos provienen de una población normal y si sustituimos σ en el intervalo (6.3.2) por la DT muestral, s , entonces la interpretación del 95 % se puede mantener si el valor que multiplica a $\frac{\sigma}{\sqrt{n}}$ (es decir, 1,96) se sustituye por una cantidad adecuada. La nueva cantidad se denomina $t_{0,025}$ y está relacionada con la distribución que se denomina distribución t de Student.

La distribución t de Student

Las **distribuciones t de Student** son distribuciones teóricas continuas que se utilizan con muchos propósitos en estadística, entre los que se incluye la construcción de intervalos de confianza. La forma exacta de una distribución t de

* Esta analogía se debe a Geoff Jowett.

Student depende de una cantidad denominada «grados de libertad», abreviadamente «gl». La Figura 6.3.2 muestra las curvas de densidad de dos distribuciones t de Student con $gl = 3$ y $gl = 10$, y también una curva normal. Una curva t es simétrica y tiene forma de campana como la curva normal, pero tiene una desviación típica mayor. A medida que gl crece, las curvas t se aproximan a la curva normal. Por tanto, la curva normal se puede ver como una curva t con gl infinito ($gl = \infty$).

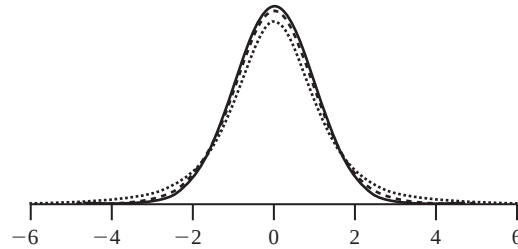


Figura 6.3.2 Dos curvas t de Student (con puntos, $gl = 3$ y con línea discontinua, $gl = 10$) y una curva normal ($gl = \infty$)

El valor $t_{0,025}$ se denomina «valor crítico al 5 % de dos colas» de la distribución t de Student y se define como el valor tal que el intervalo entre $-t_{0,025}$ y $+t_{0,025}$ contiene el 95 % del área bajo la curva, como se muestra en la Figura 6.3.3*. Es decir, el área combinada en las dos colas (por debajo de $-t_{0,025}$ y por encima de $+t_{0,025}$) es el 5 %. El área total sombreada en la Figura 6.3.3 es igual a 0,05. Nótese que el área sombreada está formada por dos «trozos» cada uno de área 0,025.

Los valores críticos de la distribución t de Student están tabulados en la Tabla 4. Los valores de $t_{0,025}$ se muestran en la columna cuyo encabezamiento es «Probabilidad de la Cola Superior 0,025». Si recorremos hacia abajo esta columna, veremos que los valores de $t_{0,025}$ decrecen a medida que gl aumenta. Para $gl = \infty$ (es decir, para la distribución normal), el valor es $t_{0,025} = 1,960$. Podemos confirmar observando la Tabla 3 que el intervalo $\pm 1,96$ (en la escala Z) contiene el 95 % del área bajo una curva normal.

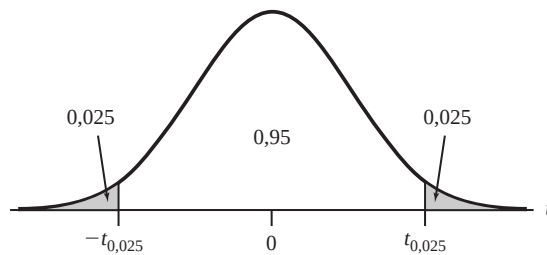


Figura 6.3.3 Definición del valor crítico $t_{0,025}$

Otras columnas de la Tabla 4 muestran otros valores críticos, que se definen de forma análoga. Por ejemplo, el intervalo $\pm t_{0,05}$ contiene el 90 % del área bajo una curva t de Student.

Intervalo de confianza para μ : método

Describiremos el método de Student para construir un intervalo de confianza para μ , basado en una muestra aleatoria de una población normal. Primero, supongamos que hemos escogido un nivel de confianza igual al 95 % (es decir, deseamos confiar un 95 %). Para construir un intervalo de confianza del 95 % para μ , calculamos los límites inferior y superior del intervalo como

$$\bar{y} - t_{0,025} ET_{\bar{y}} \quad \text{y} \quad t_{0,025} ET_{\bar{y}}$$

* En algunos textos de estadística se pueden encontrar otras notaciones, como $t_{0,05}$ o $t_{0,975}$, en vez de $t_{0,025}$.

es decir,

$$\bar{y} \pm t_{0,025} \frac{s}{\sqrt{n}}$$

donde el valor crítico $t_{0,025}$ se determina a partir de la distribución t de Student con

$$gl = n - 1$$

El siguiente ejemplo ilustra la construcción del intervalo de confianza.

Ejemplo 6.3.1 Alas de mariposas

Para los datos de la mariposa Monarch del Ejemplo 6.1.1, tenemos $n = 14$, $\bar{y} = 32,8143 \text{ cm}^2$ y $s = 2,4757 \text{ cm}^2$. La Figura 6.3.4 muestra un histograma y una gráfica de probabilidad normal de los datos. Ambas apoyan la suposición de que los datos provienen de una población normal. Tenemos 14 observaciones, por lo que el valor de gl es

$$gl = n - 1 = 14 - 1 = 13$$

De la Tabla 4 tenemos

$$t_{0,025} = 2,160$$

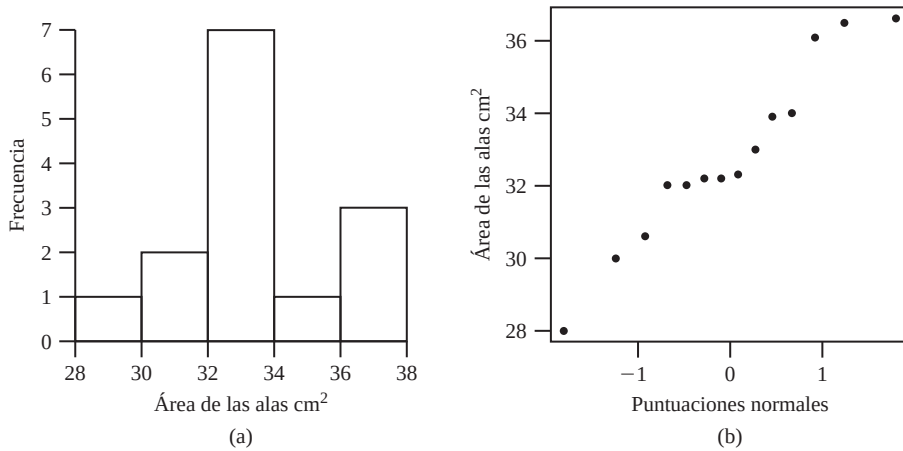


Figura 6.3.4 (a) Histograma y (b) gráfica de probabilidad normal de los datos de alas de mariposas

El intervalo de confianza del 95 % para μ es

$$\begin{aligned} 32,8143 \pm 2,160 \frac{2,4757}{\sqrt{14}} \\ 32,8143 \pm 2,160(0,6617) \\ 32,8143 \pm 1,4293 \end{aligned}$$

o, aproximadamente,

$$32,81 \pm 1,43$$

El intervalo de confianza se puede dejar de esta forma. De forma alternativa, los extremos del intervalo se pueden calcular explícitamente como

$$32,81 - 1,43 = 31,38 \quad \text{y} \quad 32,81 + 1,43 = 34,24$$

y el intervalo se puede escribir de forma más compacta de la siguiente manera

$$(31,4, 34,2)$$

o, de una forma más completa, como la siguiente «afirmación de confianza»:

$$31,4 \text{ cm}^2 < \mu < 34,2 \text{ cm}^2$$

La afirmación de confianza indica que la población de las áreas medias de las alas de las mariposas Monarch macho en la región de Océano Dunes en California está entre $31,4 \text{ cm}^2$ y $34,2 \text{ cm}^2$ con una confianza del 95 %. ■

La interpretación de la «confianza de 95 %» se comentará después del siguiente ejemplo.

Se utilizan análogamente otros coeficientes de confianza distintos del 95 %. Por ejemplo, se puede construir un intervalo de confianza del 90 % para μ utilizando $t_{0,05}$ en lugar de $t_{0,025}$ como sigue:

$$\bar{y} \pm t_{0,05} \frac{s}{\sqrt{n}}$$

Se presenta a continuación un ejemplo.

Ejemplo 6.3.2 Alas de mariposas

En la Tabla 4 encontramos que $t_{0,05} = 1,771$ con $gl = 13$. Por tanto, el intervalo de confianza del 90 % para μ en los datos de las alas de mariposas es

$$32,8143 \pm 1,771 \frac{2,4757}{\sqrt{14}}$$

$$32,8143 \pm 1,1718$$

o

$$31,6 < \mu < 34,0$$

■

Como podemos ver, la elección del nivel de confianza es algo arbitraria. Para los datos de alas de mariposas, el intervalo de confianza de 95 % es

$$32,81 \pm 1,43$$

y el intervalo de confianza del 90 % es

$$32,81 \pm 1,17$$

Por tanto, el intervalo de confianza del 90 % es más estrecho que el intervalo de confianza del 95 %. Si deseamos tener una confianza del 95 % de que nuestro intervalo contenga a μ , entonces necesitamos un intervalo más amplio que el que necesitaríamos si solo deseáramos tener un 90 % de confianza. Cuanto más alto sea el nivel de confianza, más ancho debe ser el intervalo de confianza (para un tamaño de muestra fijo; pero nótese que a medida que n aumenta, el intervalo disminuye).

Observación El valor $(n - 1)$ se denomina «grados de libertad» porque las desviaciones $y_i - \bar{y}$ deben sumar cero, y por tanto solo $(n - 1)$ de ellas son «libres» de variar. Un tamaño de muestra n proporciona solo $(n - 1)$ elementos diferentes de información sobre la variabilidad, que es, aproximadamente σ . Esto resulta particularmente claro si consideramos el caso $n = 1$. Una muestra de tamaño 1 proporciona alguna información sobre μ , pero ninguna información sobre σ , y por tanto ninguna información sobre el error de muestreo. Tiene sentido, entonces, que cuando $n = 1$, no podamos utilizar el método de la t de Student para calcular un intervalo de confianza: la desviación típica muestral no existe (véase el Ejemplo 2.6.5) y, por tanto, no hay un valor crítico con $gl = 0$. Una muestra de tamaño uno se denomina a veces «anécdota». Por ejemplo, un caso de un historial médico aislado es una anécdota. Por supuesto, un caso de un historial aislado puede contribuir en gran manera al conocimiento médico, pero no proporciona (en sí mismo) una base para juzgar lo cerca que dicho caso aislado se parece a la población en conjunto.

Intervalos de confianza y aleatoriedad

¿En qué sentido podemos tener «confianza» en un intervalo de confianza? Para responder a esta pregunta, supongamos que estamos observando una muestra aleatoria de una población normal. Consideremos, por ejemplo, un intervalo de confianza del 95 %. Una forma de interpretar el nivel de confianza (95 %) es referirse al metaestudio de muestras repetidas de la misma población. Si se construye para cada muestra un intervalo de confianza del 95 % para μ , entonces el 95 % de los intervalos de confianza contendrán a μ . Por supuesto, los datos observados en un experimento son solo *una* de las posibles muestras. Podemos esperar «confiadamente» que esta muestra sea una del 95 % afortunado, pero nunca lo sabremos.

El siguiente ejemplo muestra una visualización más concreta de la interpretación del metaestudio de un nivel de confianza.

Ejemplo 6.3.3 Espesor de cáscaras de huevo

En una cierta población grande de huevos de gallina (descrita en el Ejemplo 4.1.3), la distribución del espesor de las cáscaras es normal de media $\mu = 0,38$ mm y desviación típica $\sigma = 0,03$ mm. La Figura 6.3.5 presenta algunas muestras típicas de esta población. A la derecha se dibujan los intervalos de confianza del 95 % asociados. Los tamaños de las muestras son $n = 5$ y $n = 20$. Nótese que el segundo intervalo de confianza con $n = 5$ no contiene a μ . En la totalidad de los potenciales intervalos de confianza, el porcentaje que contendría a μ sería el 95 % para cada tamaño de la muestra. Como indica la Figura 6.3.5, muestras más grandes tienden a producir intervalos de confianza más estrechos. ■

Un intervalo de confianza se puede interpretar como una probabilidad, pero se requiere un cierto cuidado. Por ejemplo, si consideramos intervalos de confianza del 95 %, entonces la siguiente afirmación es correcta:

$$\Pr\{\text{la siguiente muestra nos proporcionará un intervalo de confianza que contenga a } \mu\} = 0,95$$

Sin embargo, hay que darse cuenta de que es *el intervalo de confianza* lo que es aleatorio en esta afirmación, y no es correcto sustituirlo con su valor obtenido a partir de los datos. Así, por ejemplo, en el Ejemplo 6.3.1 obtuvimos que el intervalo de confianza del 95 % para la media de las alas de mariposas es

$$31,4 \text{ cm}^2 < \mu < 34,2 \text{ cm}^2 \quad (6.3.3)$$

No obstante, *no* es correcto decir que

$$\Pr\{31,4 \text{ cm}^2 < \mu < 34,2 \text{ cm}^2\} = 0,95$$

porque esta afirmación no contiene ningún elemento aleatorio: o μ está entre 20,6 y 22,1 o no lo está. Si $\mu = 32$, entonces $\Pr\{31,4 \text{ cm}^2 < \mu < 34,2 \text{ cm}^2\} = \Pr\{31,4 \text{ cm}^2 < 32 < 34,2 \text{ cm}^2\} = 1$ (no 0,95). La siguiente analogía puede ayudar a clarificar este punto.

Supongamos que Y representa el número de puntos que se muestran cuando se lanza un dado equilibrado. Entonces

$$\Pr\{Y = 2\} = \frac{1}{6}$$

Por otra parte, si ahora lanzamos el dado y observamos 5 puntos, obviamente *no* es correcto sustituir este «dato» en la afirmación sobre la probabilidad, y concluir que

$$\Pr\{5 = 2\} = \frac{1}{6}^*$$

Como indica la precedente exposición, el nivel de confianza (por ejemplo, del 95 %) es una propiedad del *método* más bien que de un intervalo particular. Una afirmación individual, como la (6.3.3) es o verdadera o falsa, pero a largo plazo, si un investigador construye intervalos de confianza del 95 % en varios experimentos, y cada vez se produce una afirmación como la (6.3.3), entonces el 95 % de las afirmaciones serán ciertas.

* Incluso aunque el dado ruede debajo de una silla y no pudiéramos ver inmediatamente que la cara superior tiene cinco puntos, sería incorrecto (dada nuestra definición de probabilidad) decir «la probabilidad de que la cara superior del dado muestre 2 puntos es 1/6».

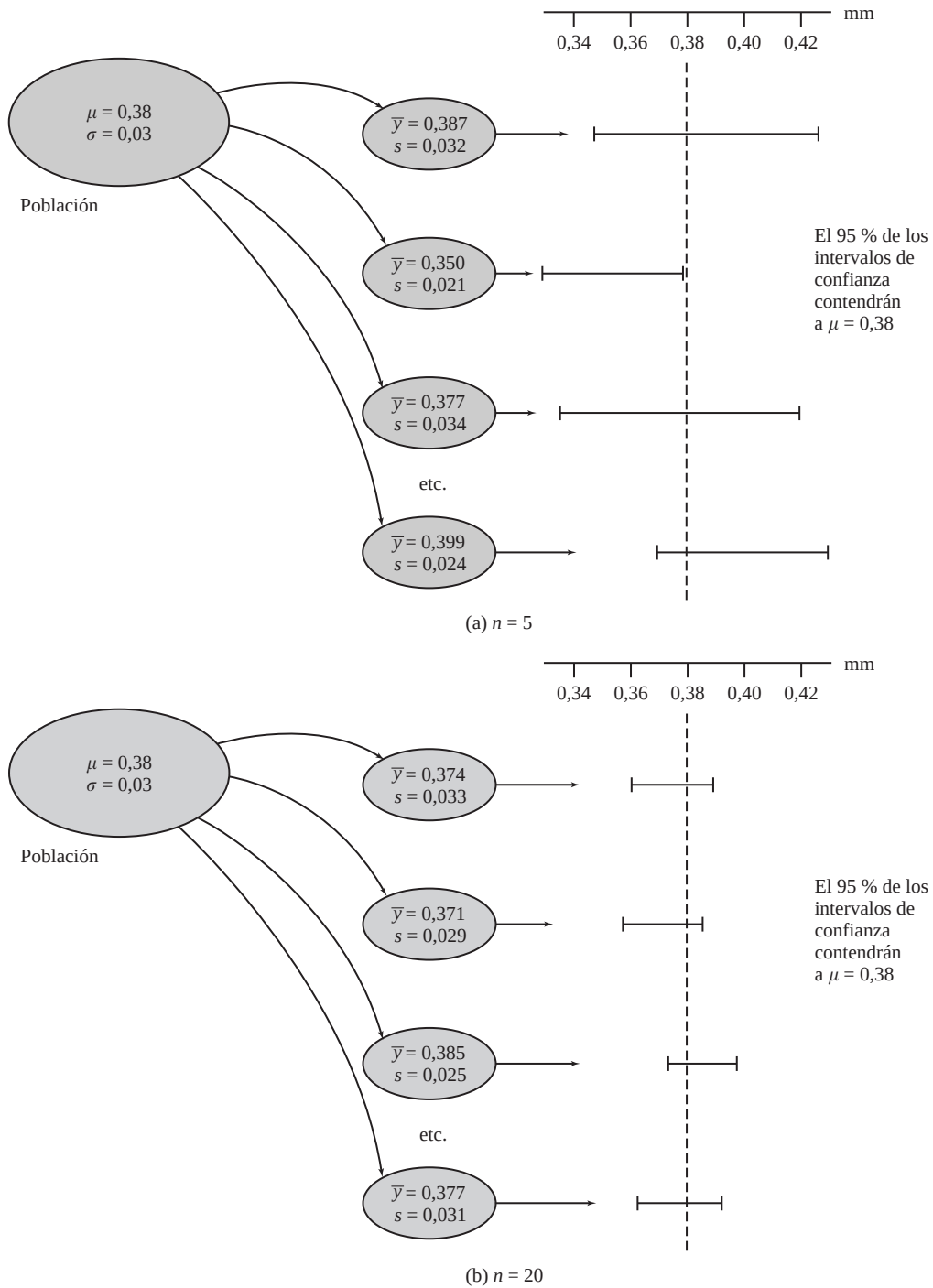


Figura 6.3.5 Intervalos de confianza para el espesor medio de las cáscaras de huevo

Interpretación de un intervalo de confianza

Ejemplo 6.3.4 Densidad mineral del hueso

Una densidad mineral del hueso baja produce a menudo fracturas de cadera en personas mayores. En un experimento para evaluar la efectividad de una terapia de sustitución hormonal, unos investigadores proporcionaron estrógeno equino conjugado (EEC) a una muestra de 94 mujeres entre 45 y 64 años¹⁰. Tras tomar la medicación durante 36 meses, se midió la densidad mineral del hueso en las 94 mujeres. La densidad media fue de $0,878 \text{ g/cm}^2$, con una desviación típica de $0,126 \text{ g/cm}^2$.

El error típico de la media es por tanto $\frac{0,126}{\sqrt{94}} = 0,013$. No está claro que la distribución de la densidad mineral del hueso tenga una distribución normal, pero como veremos en la Sección 6.5, cuando el tamaño de la muestra es grande, la condición de normalidad no es crucial. Hay 94 observaciones, por lo que hay 93 grados de libertad. Para obtener el multiplicador t para un intervalo de confianza del 95 %, utilizaremos 100 grados de libertad (ya que la Tabla 4 no muestra el caso de 93 grados de libertad). El multiplicador t es $t_{0,025} = 1,984$. Un intervalo de confianza del 95 % para μ es

$$0,878 \pm 1,984(0,013)$$

o, aproximadamente,

$$0,878 \pm 0,026$$

o

$$(0,852, 0,904)^*$$

Por tanto, tenemos una confianza del 95 % de que la densidad mineral del hueso de la cadera de las mujeres con edades entre 45 y 64 años que tomaron EEC durante 36 meses esté entre $0,852 \text{ g/cm}^2$ y $0,904 \text{ g/cm}^2$. ■

Ejemplo 6.3.5 Semillas por fruto

El número de semillas por fruto de la planta de agua dulce *Vallisneria Americana* varía considerablemente de un fruto a otro. Un investigador¹¹ tomó una muestra aleatoria de 12 frutos y encontró que el número medio de semillas era de 320, con una desviación típica de 125. El investigador esperaba que el número de semillas siguiera, al menos aproximadamente, una distribución normal. La Figura 6.3.6 muestra una gráfica de probabilidad normal de los datos. Esta gráfica apoya el uso de un modelo de distribución normal para estos datos.

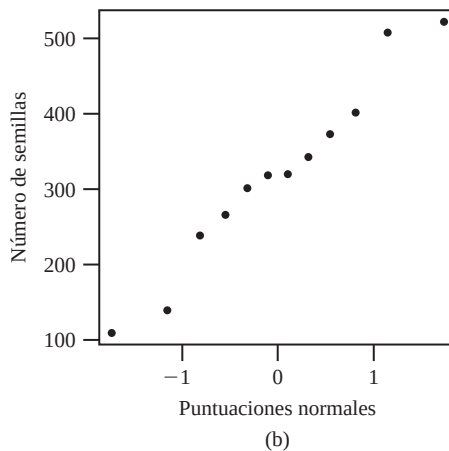


Figura 6.3.6 Gráfica de probabilidad normal de las semillas por fruto de la *Vallisneria Americana*

* Si utilizamos un computador para calcular el intervalo de confianza, obtenemos (0,8522, 0,9038). Hay muy poca diferencia entre los multiplicadores t para 93 y 100 grados de libertad.

El error típico de la media es $\frac{125}{\sqrt{12}} = 36$. Hay 11 grados de libertad. El multiplicador t para un intervalo de confianza del 90 % es $t_{0,05} = 1,796$. Un intervalo de confianza del 90 % para μ es

$$320 \pm 1,796(36)$$

o, aproximadamente,

$$320 \pm 65$$

o

$$(255, 385)$$

Por tanto, *tenemos una confianza del 90 % de que el número medio (poblacional) de semillas por fruto de la Vallisneria Americana esté entre 255 y 385.* ■

Relación con la distribución muestral de $\bar{Y}$

En este punto puede ser útil mirar atrás y ver cómo se relaciona el intervalo de confianza para μ con la distribución muestral de $\bar{Y}$. Recuerdese, de la Sección 5.3, que la media de la distribución muestral es μ y su desviación típica es $\frac{\sigma}{\sqrt{n}}$.

La Figura 6.3.7 muestra una media muestral particular ($\bar{y}$) y su intervalo de confianza del 95 % para μ asociado, superpuesto a la distribución muestral de $\bar{Y}$. Nótese que el intervalo de confianza concreto contiene a μ . Esto ocurrirá para el 95 % de las muestras.

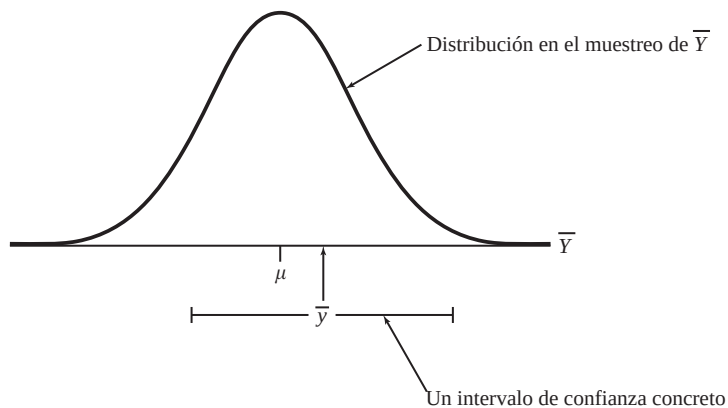


Figura 6.3.7 Relación entre un intervalo de confianza concreto para μ y la distribución muestral de $\bar{Y}$

Intervalos de confianza unilaterales

La mayoría de los intervalos de confianza son de la forma «estimación $\pm$ margen de error», y se conocen como intervalos de confianza bilaterales. Sin embargo, es posible construir un intervalo de confianza unilateral, que será lo apropiado cuando solo es de interés una cota inferior o una cota superior. Los dos ejemplos siguientes ilustran los intervalos de confianza unilaterales del 90 % y del 95 %.

Ejemplo 6.3.6 Semillas por fruto. Unilateral al 90 %

Consideremos los datos de semillas del Ejemplo 6.3.5, que se utilizan para estimar el número de semillas por fruto de la *Vallisneria Americana*. Podría ser que deseáramos una cota inferior de μ , la media poblacional, pero que no nos importara lo grande que pudiera ser μ . Mientras que un intervalo de confianza bilateral del 90 % se basa en capturar el 90 % central de una distribución t y, por tanto, usa los multiplicadores t correspondientes a $\pm t_{0,05}$, un

intervalo de confianza unilateral del 90 % (inferior) utiliza el hecho de que $\Pr(-t_{0,10} < t < \infty) = 0,90$. Por tanto, el límite inferior del intervalo de confianza es $\bar{y} - t_{0,10} ET_{\bar{y}}$ y el límite superior del intervalo es infinito. En este caso, con 11 grados de libertad, el multiplicador t es $t_{11,0,10} = 1,363$, y obtenemos

$$320 - 1,363(36) = 320 - 49 = 271$$

como límite inferior. El intervalo resultante es $(271, \infty)$. Por tanto, *tenemos una confianza del 90 % de que el número medio (poblacional) de semillas por fruto de la Vallisneria Americana sea al menos de 271.* ■

Ejemplo 6.3.7 Semillas por fruto. Unilateral al 95 %

Un intervalo de confianza unilateral del 95 % se construye de la misma forma que un intervalo de confianza unilateral del 90 %, pero con un multiplicador t diferente. Para los datos de semillas de *Vallisneria Americana* tenemos $t_{11,0,05} = 1,796$ y, por tanto,

$$320 - 1,796(36) = 320 - 65 = 255$$

como límite inferior. El intervalo resultante es $(255, \infty)$. Por tanto, *tenemos una confianza del 95 % de que el número medio (poblacional) de semillas por fruto de la Vallisneria Americana sea al menos de 255.* ■

Ejercicios 6.3.1-6.3.20

6.3.1 (Ejercicio de muestreo) En referencia al Ejercicio 5.3.1, utilice su muestra de cinco longitudes de elipses para construir un intervalo de confianza del 80 % para μ , utilizando la fórmula $\bar{y} \pm (1,533)s/\sqrt{n}$.

6.3.2 (Ejercicio de muestreo) En referencia al Ejercicio 5.3.3, utilice su muestra de 20 longitudes de elipses para construir un intervalo de confianza del 80 % para μ , utilizando la fórmula $\bar{y} \pm (1,328)s/\sqrt{n}$.

6.3.3 Como parte de un estudio del desarrollo de la glándula timo, unos investigadores pesaron las glándulas de cinco embriones de pollo tras 14 días de incubación. Los pesos del timo (mg) fueron los siguientes¹²:

29,6 21,5 28,0 34,6 44,9

Para esos datos, la media es 31,7 y la desviación típica 8,7.

- Calcule el error típico de la media.
- Construya un intervalo de confianza del 90 % para la media poblacional.

6.3.4 Considere los datos del Ejercicio 6.3.3.

- Construya un intervalo de confianza del 95 % para la media poblacional.
- Interprete el intervalo de confianza obtenido en el apartado (a). Es decir, explique lo que significan los números del intervalo (véanse los Ejemplos 6.3.4 y 6.3.5).

6.3.5 Se inyectó el antibiótico Gentamicina a seis ovejas sanas de tres años, con una dosis de 10 mg/kg de peso corporal. Sus concentraciones en suero ($\mu\text{g/ml}$) de Gentamicina 1,5 horas después de la inyección fueron las siguientes¹³:

33 26 34 31 23 25

Para esos datos, la media es de 28,7 y la desviación típica es 4,6.

- Construya un intervalo de confianza del 95 % para la media poblacional.

- Interprete el intervalo de confianza obtenido en el apartado (a) (véase el Ejemplo 6.1.1).

- El intervalo definido en el apartado (a) contiene casi todas las observaciones. ¿Será esto cierto en general para un intervalo de confianza del 95 %? Explique su respuesta.

6.3.6 Un zoólogo midió la longitud de la cola de 86 individuos, todos en el grupo de un año de edad, del ratón ciervo *Peromyscus*. La longitud media fue de 60,43 mm y la desviación típica fue de 3,06 mm. Un intervalo de confianza del 95 % para la media es (59,77, 61,09).

- Verdadero o falso (e indique por qué): tenemos una confianza del 95 % de que el promedio de las longitudes de cola de los 86 individuos de la muestra esté entre 59,77 mm y 61,09 mm.
- Verdadero o falso (e indique por qué): tenemos una confianza del 95 % de que el promedio de las longitudes de cola de todos los individuos de la población esté entre 59,77 mm y 61,09 mm.

6.3.7 En referencia al Ejercicio 6.3.6.

- Sin realizar los cálculos, ¿sería un intervalo de confianza del 80 % para la longitud media de la cola más ancho, más estrecho o aproximadamente igual? Explique su respuesta.
- Sin realizar los cálculos, si se midieran 500 ratones en vez de 86, ¿sería el intervalo de confianza del 95 % indicado en el Ejercicio 6.3.6, más ancho, más estrecho o aproximadamente igual? Explique su respuesta.

6.3.8 Unos investigadores midieron la densidad mineral del hueso de la columna vertebral de 94 mujeres después de que tomaran el medicamento EEC. (Véase el Ejemplo 6.3.4, que trataba de la densidad mineral del hueso de la cadera).

La media fue de 1,016 g/cm² y la desviación típica de 0,155 g/cm². Un intervalo de confianza para la media es (0,984, 1,048).

- (a) Verdadero o falso (e indique por qué): el 95 % de las medidas de densidad mineral del hueso de la muestra están entre 0,984 y 1,048.
- (b) Verdadero o falso (e indique por qué): el 95 % de las medidas de densidad mineral del hueso de la población están entre 0,984 y 1,048.

6.3.9 En el estudio descrito en el Ejemplo 6.3.4 había un grupo de control. Las 124 mujeres del grupo de control recibieron un placebo, en vez de una medicación activa. Al final del estudio tuvieron una densidad mineral media del hueso de 0,840 g/cm². A continuación se muestran tres intervalos de confianza: uno es un intervalo de confianza del 90 %, otro del 85 % y otro del 80 %. Sin realizar cálculos, indique qué intervalo se corresponde con cada nivel de confianza y explique cómo lo ha determinado.

Nivel de confianza:

90 % 85 % 80 %

Intervalos (en orden mezclado):

(0,826, 0,854) (0,824, 0,856) (0,822, 0,858)

6.3.10 La beta-endorfina humana (BEH) es una hormona segregada por la glándula pituitaria bajo condiciones de estrés. Un investigador realizó un estudio para investigar si un programa de ejercicio regular podría afectar a las concentraciones en reposo (sin estrés) de BEH en la sangre. Midió los niveles de BEH en sangre, en enero y de nuevo en mayo, de 10 participantes en un programa de ejercicio físico. Los resultados se muestran en la tabla¹⁴.

- (a) Construya un intervalo de confianza del 95 % para la diferencia de medias poblacionales de niveles de BEH entre enero y mayo. (Sugerencia: solo necesita utilizar los valores de la columna de la derecha).

Participante	Nivel de BEH (pg/ml)		
	Enero	Mayo	Diferencia
1	42	22	20
2	47	29	18
3	37	9	28
4	9	9	0
5	33	26	7
6	70	36	34
7	54	38	16
8	27	32	-5
9	41	33	8
10	18	14	4
Media	37,8	24,8	13,0
DT	17,6	10,9	12,4

- (b) Interprete el intervalo de confianza del apartado (a). Es decir, explique lo que le indica el intervalo sobre los niveles de BEH. (Véanse los Ejemplos 6.3.4 y 6.3.5).
- (c) Utilizando su intervalo para apoyar su respuesta, ¿existe evidencia de los niveles de BEH son menores en mayo que en enero? (Sugerencia: ¿incluye su intervalo el valor cero?)

6.3.11 Considere los datos del Ejercicio 6.3.10. Si el tamaño de la muestra es pequeño, como en este caso, entonces para que un intervalo de confianza basado en la distribución t de Student sea válido, los datos deben provenir de una población distribuida normalmente. ¿Es razonable pensar que la diferencia de niveles de BEH está distribuida normalmente? ¿Cómo lo sabe?

6.3.12 La invertasa es una enzima que puede ayudar a la germinación de las esporas del hongo *Colletotrichum graminicola*. Un botánico incubó especímenes del tejido del hongo en placas de Petri y posteriormente ensayó la actividad de la invertasa sobre los tejidos. Los valores de actividad específica para nueve placas de Petri incubadas con una humedad relativa del 90 % durante 24 horas se resumen como sigue¹⁵:

Media = 5.111 unidades DT = 818 unidades

- (a) Suponga que los datos son una muestra aleatoria de una población normal. Construya un intervalo de confianza del 95 % para la actividad media de la invertasa bajo esas condiciones experimentales.
- (b) Interprete el intervalo de confianza obtenido en el apartado (a). Es decir, explique lo que significan los números del intervalo. (Véanse los Ejemplos 6.3.4 y 6.3.5).
- (c) Si tuviera los datos originales, ¿cómo podría comprobar la condición de que los datos provienen de una población normal?

6.3.13 Como parte de un estudio del tratamiento de la anemia en el ganado, unos investigadores midieron la concentración de selenio en la sangre de 36 vacas que habían recibido una dieta suplementaria de selenio (2 mg/día) durante un año. Las vacas eran de la misma raza (*Santa Gertrudis*) y habían tenido su primer ternero durante el año. La concentración media de selenio fue de 6,21 μ g/dl y su desviación típica fue de 1,84 μ g/dl¹⁶. Construya un intervalo de confianza del 95 % para la media poblacional.

6.3.14 En un estudio sobre el desarrollo de las larvas de la polilla de la manzana copetuda (*Platynota idaeusalis*), un entomólogo midió la anchura de la cabeza de 50 larvas. Todas ellas se habían criado bajo idénticas condiciones y habían mudado seis veces. La anchura media de la cabeza fue de 1,20 mm y la desviación típica fue de 0,14 mm. Construya un intervalo de confianza del 90 % para la media poblacional¹⁷.

6.3.15 En un estudio sobre el efecto del consumo de aluminio en el desarrollo mental de los bebés, se suministró a 92 bebés que habían nacido prematuramente una solución especial baja en aluminio por vía intravenosa¹⁸. A la edad de

18 meses se midió el desarrollo neurológico de los bebés mediante el Índice de Desarrollo Mental de Bayley. (El Índice de Desarrollo Mental de Bayley es similar a la puntuación de CI, y el promedio de la población general es 100). Un intervalo de confianza del 95 % para la media es (93,8, 102,1).

- (a) Interprete este intervalo. Es decir, ¿qué nos dice el intervalo sobre el desarrollo neurológico en la población de bebés nacidos prematuramente que recibieron soluciones por vía intravenosa?
- (b) ¿Indica este intervalo que el CI medio de la población muestreada está por debajo del promedio de 100 de la población general?

6.3.16 Un grupo de 101 pacientes en la etapa final de una enfermedad renal fueron tratados con el medicamento epoetin¹⁹. El nivel medio de hemoglobina de los pacientes fue de 10,3 (g/dl), con una DT de 0,9. Construya un intervalo de confianza del 95 % para la media poblacional.

6.3.17 En la Tabla 4 encontramos que $t_{0,025} = 1,960$ cuando $gl = \infty$. Demuestre cómo se puede verificar este valor utilizando la Tabla 3.

6.3.18 Utilice la Tabla 3 para verificar el valor de $t_{0,0025}$ cuando $gl = \infty$. (No intente interpolar en la Tabla 4).

6.3.19 Los datos se resumen a menudo en este formato: $\bar{y} \pm ET$. Suponga que este intervalo se interpreta como un intervalo de confianza. Si el tamaño de la muestra es grande, ¿cuál sería el nivel de confianza de ese intervalo? Es decir, ¿cuál es la probabilidad de que un intervalo calculado como

$$\bar{y} \pm (1,00)ET$$

contenga realmente a la media poblacional? [*Sugerencia:* recuerde que el nivel de confianza del intervalo $\bar{y} \pm (1,96)ET$ es del 95 %].

6.3.20 (Continuación del Ejercicio 6.3.19)

- (a) Si el tamaño de la muestra es pequeño pero la distribución poblacional es normal, ¿es el nivel de confianza del intervalo $\bar{y} \pm ET$ mayor o menor que la respuesta al Ejercicio 6.3.19? Explique su contestación.
- (b) ¿Cómo resulta afectada la respuesta al Ejercicio 6.3.19 si la distribución poblacional de Y no es aproximadamente normal?

6.4 Planificación de un estudio para estimar μ

Antes de recoger datos para un estudio de investigación, es sensato considerar previamente si las estimaciones generadas a partir de los datos serán suficientemente precisas. Puede ser de verdad doloroso descubrir tras un largo y costoso estudio que los errores típicos son tan grandes que las cuestiones principales planteadas en el estudio no se pueden resolver.

La precisión con la que se puede estimar una media poblacional está determinada por dos factores: (1) la variabilidad poblacional de la variable observada, Y , y (2) el tamaño de la muestra.

En algunas situaciones la variabilidad de Y no se puede, o quizá no se debería, reducir. Por ejemplo un ecólogo de la fauna y flora puede desear realizar un estudio de campo de una población natural de peces. La heterogeneidad de la población no es controlable y de hecho es un objetivo propio de la investigación. Como otro ejemplo, en una investigación médica, además de conocer la respuesta media a un tratamiento, puede ser importante saber cómo varía la respuesta de un paciente a otro y, por tanto, puede resultar inapropiado utilizar un grupo de pacientes demasiado homogéneo.

Por otra parte, a menudo es apropiado, especialmente en estudios comparativos, reducir la variabilidad de Y manteniendo las condiciones *externas* tan constantes como sea posible. Por ejemplo, las medidas fisiológicas se pueden tomar a una hora fija del día. Los tejidos se pueden mantener a temperatura constante. Todos los animales utilizados en un experimento deben ser de la misma edad.

Supongamos, entonces, que se han hecho planes para reducir la variabilidad de Y tanto como sea posible, o deseable. ¿Qué tamaño de la muestra será suficiente para conseguir un grado de precisión deseado en la estimación de la media poblacional? Si utilizamos el error típico como nuestra medida de precisión, entonces esta pregunta se puede plantear de una forma directa. Recuérdesse que el ET se define como

$$ET_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

Para decidir sobre un valor de n , se debe (1) especificar el valor de ET que se considera deseable y (2) tener disponible un valor preliminar de la DT, bien a partir de un estudio piloto, de la experiencia previa o de la literatura científica. El tamaño de la muestra requerido se determina entonces a partir de la siguiente ecuación:

$$ET \text{ deseado} = \frac{\text{Valor preliminar de la DT}}{\sqrt{n}}$$

El siguiente ejemplo ilustra el uso de esta ecuación.

Ejemplo 6.4.1 Alas de mariposas

Los datos del Ejemplo 6.1.1 de las alas de mariposas produjeron los siguientes estadísticos resumen:

$$\bar{y} = 32,81 \text{ cm}^2$$

$$s = 2,48 \text{ cm}^2$$

$$ET = 0,66 \text{ cm}^2$$

Supongamos que la investigadora planea ahora un nuevo estudio de mariposas y ha decidido que sería deseable que el ET no fuera superior a $0,4 \text{ cm}^2$. Como valor preliminar de la DT utiliza el valor del estudio anterior, concretamente $2,48 \text{ cm}^2$. Por tanto, el valor de n deseado debe cumplir la siguiente relación:

$$ET = \frac{2,48}{\sqrt{n}} \leq 0,4$$

Esta ecuación se puede resolver fácilmente obteniéndose $n \geq 38,4$. Como no se pueden tener 38,4 mariposas, el nuevo estudio debe incluir al menos 39 mariposas. ■

Podemos preguntarnos cómo llegaría un investigador a un valor como $0,4 \text{ cm}^2$ para el ET deseado. Ese valor se determina considerando cuánto error estamos dispuestos a tolerar al estimar μ . Por ejemplo, supongamos que la investigadora del Ejemplo 6.4.1 hubiera decidido que le gustaría poder estimar la media poblacional, μ , con una precisión de $\pm 0,8$ con una confianza del 95 %. Es decir, desearía que su intervalo de confianza del 95 % para μ fuera $\bar{y} \pm 0,8$. La «parte $\pm$ del intervalo de confianza, que se denomina algunas veces **margen de error con un 95 % de confianza**, es $t_{0,025} \times ET$. El valor preciso de $t_{0,025}$ depende de los grados de libertad, pero en general, $t_{0,025}$ vale aproximadamente 2. Por tanto, la investigadora desea que $2 \times ET$ no sea mayor que 0,8. Esto significa que el error no debería ser mayor que $0,4 \text{ cm}^2$.

En estudios comparativos, la consideración principal es generalmente el tamaño de los efectos anticipados del tratamiento. Por ejemplo, si se planea comparar dos grupos experimentales de poblaciones distintas, el ET anticipado de cada población o grupo experimental debería ser sustancialmente menor que (preferiblemente menor que un cuarto de) la diferencia anticipada de las medias de los dos grupos*. Por tanto, la investigadora de mariposas del Ejemplo 6.4.1 podría llegar al valor de $0,4 \text{ cm}^2$ si estuviera planeando comparar mariposas Monarch macho y hembra y esperara que las áreas de las alas de los sexos difirieran (en promedio) aproximadamente $1,6 \text{ cm}^2$. Entonces debería capturar 39 mariposas macho y 39 mariposas hembra.

Para ver cómo el valor de n requerido depende de la precisión especificada, suponga que la investigadora de mariposas especificara que el ET deseado fuera de $0,2 \text{ cm}^2$ en vez de $0,4 \text{ cm}^2$. Entonces la relación sería

$$ET = \frac{2,48}{\sqrt{n}} \leq 0,2$$

lo que resulta en $n = 153,76$, por lo que debería capturar 154 mariposas de cada sexo. Por tanto, doblar la precisión (disminuyendo a la mitad el ET) no requiere dos veces, sino cuatro veces más observaciones. Este fenómeno de «rendimientos decrecientes» se debe a la raíz cuadrada en la fórmula del ET.

Ejercicios 6.4.1-6.4.5

6.4.1 Se está planeando un experimento para comparar los efectos de varias dietas en la ganancia de peso del ganado, medida durante un periodo de prueba de 140 días²⁰. Para tener suficiente precisión al comparar las dietas, se desea que

el error típico de la media de cada dieta no debería exceder los 5 kg.

(a) Si se tiene un valor preliminar de la desviación típica de la ganancia de peso de aproximadamente 20 kg en cual-

* Esto es una guía aproximada para obtener una sensibilidad adecuada para discriminar tratamiento. Esa sensibilidad, denominada técnicamente *potencia*, se estudia en el Capítulo 7.

quiera de las dietas, ¿qué número de cabezas de ganado debería seguir cada dieta para conseguir un error típico suficientemente pequeño?

- (b) Si se dobla el valor preliminar de la desviación típica, hasta 40 kg, se doblaría el número requerido de cabezas de ganado? Explique su respuesta.

6.4.2 Un investigador médico propone estimar el nivel medio de colesterol en suero de una cierta población de hombres de mediana edad, basándose en una muestra aleatoria de la población. Se asesora con un estadístico. La subsiguiente discusión revela que el investigador desea estimar la media poblacional con una precisión de ± 6 mg/dl o menos, con un 95 % de confianza. Por tanto, el error típico de la media debería ser de 3 mg/dl o menos. Además, el investigador cree que la desviación típica del colesterol en suero de la población está probablemente sobre los 40 mg/dl²¹. ¿Qué tamaño de muestra debería tomar el investigador?

6.4.3 Una fisióloga de plantas planea medir las longitudes del tallo de plantas de soja tras tres semanas de crecimiento utilizando un nuevo fertilizante. Los experimentos previos sugieren que la desviación típica de la longitud del tallo está sobre los 1,2 cm²². Utilizando esto como un valor preliminar de σ , determine cuántas plantas de soja debería tener la investigadora si desea que el error típico de la media del grupo no sea superior a 0,2 cm.

6.4.4 Suponga que está planeando un experimento para probar los efectos de varias dietas en la ganancia de peso de pavos jóvenes. La variable observada será Y = ganancia de peso en tres semanas (medida en un período que empieza una semana después del nacimiento y termina tres semanas después). Los experimentos previos sugieren que la desviación típica de Y para una dieta estándar es aproximadamente de 80 g²³. Utilizando esto como un valor inicial de σ , determine cuántos pavos debería tener en el grupo de tratamiento, si desea que el error típico de la media del grupo no sea mayor que

- (a) 20 g
(b) 15 g

6.4.5 Una investigadora planea comparar los efectos de dos tipos diferentes de luz en el crecimiento de plantas de judía. Espera que las medias de los dos grupos difieran aproximadamente 1 pulgada y que en cada grupo la desviación típica del crecimiento de las plantas esté alrededor de 1,5 pulgadas. Considere la guía de que el ET anticipado de cada grupo experimental no debe ser mayor que la cuarta parte de la diferencia anticipada entre las dos medias de los grupos. ¿Qué tamaño debería tener la muestra (en cada grupo) para cumplir este requisito?

6.5 Condiciones de validez de los métodos de estimación

Dada cualquier muestra de datos cuantitativos, podemos utilizar los métodos de este capítulo para calcular la media, su error típico, y varios intervalos de confianza. De hecho, los computadores pueden facilitar la realización de estas tareas. Sin embargo, las *interpretaciones* que hemos dado para esa descripción de los datos solo son válidas bajo ciertas condiciones.

Condiciones para la validez de la fórmula del ET

Primero, la noción de considerar la media muestral como un estimador de la media poblacional requiere que los datos se vean «como si» hubieran sido generados mediante el muestreo aleatorio de alguna población. En tanto en cuanto esto no sea posible, cualquier inferencia más allá de los datos reales será cuestionable. El ejemplo siguiente ilustra esta dificultad.

Ejemplo 6.5.1 Marihuana e inteligencia

Diez personas que consumían marihuana de forma intensiva resultaron ser muy inteligentes: su CI medio fue de 128,4, mientras que se sabe que la media de la población general es de 100. Las 10 personas pertenecían a un grupo religioso que utilizaba la marihuana con propósitos rituales. Como su decisión de unirse al grupo podría muy bien estar relacionada con su inteligencia, no es claro que los 10 se puedan considerar (con respecto al CI) como una muestra aleatoria de una población particular, y por tanto no existe base para pensar que la media muestral (128,4) sea un estimador de la media poblacional de una población particular (como, por ejemplo, todos los que consumen marihuana de forma intensiva). Cualquier inferencia sobre el *efecto* de la marihuana en el CI sería incluso más inverosímil, especialmente porque no hay disponibles datos de los CI de las 10 personas *antes* de que empezaran a consumir marihuana²⁴.

Segundo, el uso de la fórmula del error típico $ET = s/\sqrt{n}$ requiere dos condiciones adicionales:

1. El tamaño de la población debe ser grande comparado con el tamaño de la muestra. Este requisito raramente es un problema en las Ciencias de la Vida. La muestra puede llegar a ser un 5 % de la población sin invalidar seriamente la fórmula del ET*.
2. Las observaciones deben ser independientes entre sí. Este requisito significa que las n observaciones proporcionan realmente n piezas diferentes de información sobre la población.

Los datos a menudo no cumplen el requisito de independencia si el experimento o el sistema de muestreo tienen una **estructura jerárquica**, en la que las unidades observacionales están «anidadas» dentro de las unidades de muestreo, como se ilustra en el siguiente ejemplo.

Ejemplo 6.5.2 Anatomía canina

El músculo coccígeo es un músculo bilateral en la región pélvica del perro. Como parte de un estudio anatómico, se pesaron el lado derecho y el lado izquierdo del músculo coccígeo en 21 perras. Por tanto, había $2 \times 21 = 42$ observaciones, pero solo 21 unidades escogidas de la población de interés (perras). Debido a la simetría del músculo coccígeo, la información contenida en los lados derecho e izquierdo es muy redundante, de forma que los datos no contienen 42, sino solo 21 piezas de información sobre el músculo coccígeo de las perras. Sería por tanto incorrecto aplicar la fórmula del ET como si los datos fueran de una muestra de tamaño $n = 42$. La estructura jerárquica de los datos se indica en la Figura 6.5.1²⁵. ■

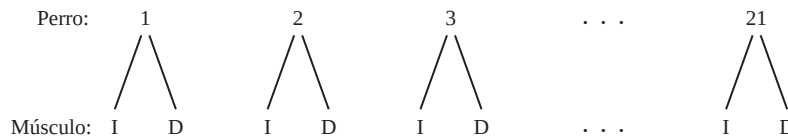


Figura 6.5.1 Estructura jerárquica de los datos del Ejemplo 6.5.2

Las estructuras de datos jerárquicas son bastante comunes en las Ciencias de la Vida. Por ejemplo, las observaciones realizadas en 90 células nerviosas pueden provenir solo de tres gatos diferentes; 80 granos de trigo que provienen de la misma espiga; 60 ratones jóvenes procedentes de solo 10 camadas. Un ejemplo particularmente claro de observaciones no independientes es la repetición de medidas sobre un mismo individuo. Por ejemplo si un médico realiza medidas triples de la presión sanguínea en 10 pacientes, claramente no dispone de 30 observaciones independientes. En algunas situaciones es obvio un tratamiento correcto de los datos jerárquicos. Por ejemplo, las medidas triples de la presión sanguínea se podrían promediar para dar un único valor de cada paciente. Sin embargo, en otras situaciones, la falta de independencia puede ser más sutil. Por ejemplo, supongamos que se incluyen 60 ratones jóvenes procedentes de 10 camadas en un experimento para comparar dos dietas. Entonces, la elección de un análisis correcto depende del *diseño* del experimento, en aspectos tales como si las dietas se aplicaron a los ratones jóvenes o a las madres, y cómo se asignaron los animales a las dos dietas.

Algunas veces aparece variación en varios niveles jerárquicos diferentes de un experimento, y puede ser un desafío evitarlos todos, y particularmente, identificar correctamente el valor de n . El Ejemplo 6.5.3 ilustra este asunto.

Ejemplo 6.5.3 Germinación de esporas

En un estudio del hongo que causa la enfermedad de la antracnosis en el maíz, el interés se centró en la supervivencia de las esporas del hongo²⁶. Se almacenaron en cámaras grupos de esporas, todas ellas preparadas a partir de un único cultivo del hongo, bajo diversas condiciones ambientales y se probó su capacidad de germinación de la siguiente forma. Cada grupo de esporas se suspendió en agua y después se cultivó en agar en una placa de Petri. Se

* Si el tamaño de la muestra, n , es una fracción sustancial del tamaño de la población, N , debe aplicarse el «factor de población finita». Este factor es $\sqrt{\frac{N-n}{N-1}}$. El error típico de la media se expresa entonces como $\frac{s}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$.

extrajeron tres trozos de 3 mm de diámetro de cada placa de Petri y se incubaron a 12 °C durante 12 horas. Cada trozo se examinó después al microscopio para buscar las esporas germinadas y las no germinadas. Las condiciones ambientales de almacenamiento (los «tratamientos») incluían los siguientes:

T_1 : Almacenamiento durante una semana con una humedad relativa del 70 %

T_2 : Almacenamiento durante una semana con una humedad relativa del 60 %

T_3 : Almacenamiento durante dos semanas con una humedad relativa del 60 %

y así sucesivamente.

En total había 43 tratamientos.

El diseño del experimento se muestra esquemáticamente en la Figura 6.5.2. Había 129 grupos de esporas, que se asignaron aleatoriamente a los 43 tratamientos, a razón de tres grupos por tratamiento. Cada grupo de esporas se ubicó en una placa de Petri, y de cada placa de Petri se extrajeron 10 trozos.

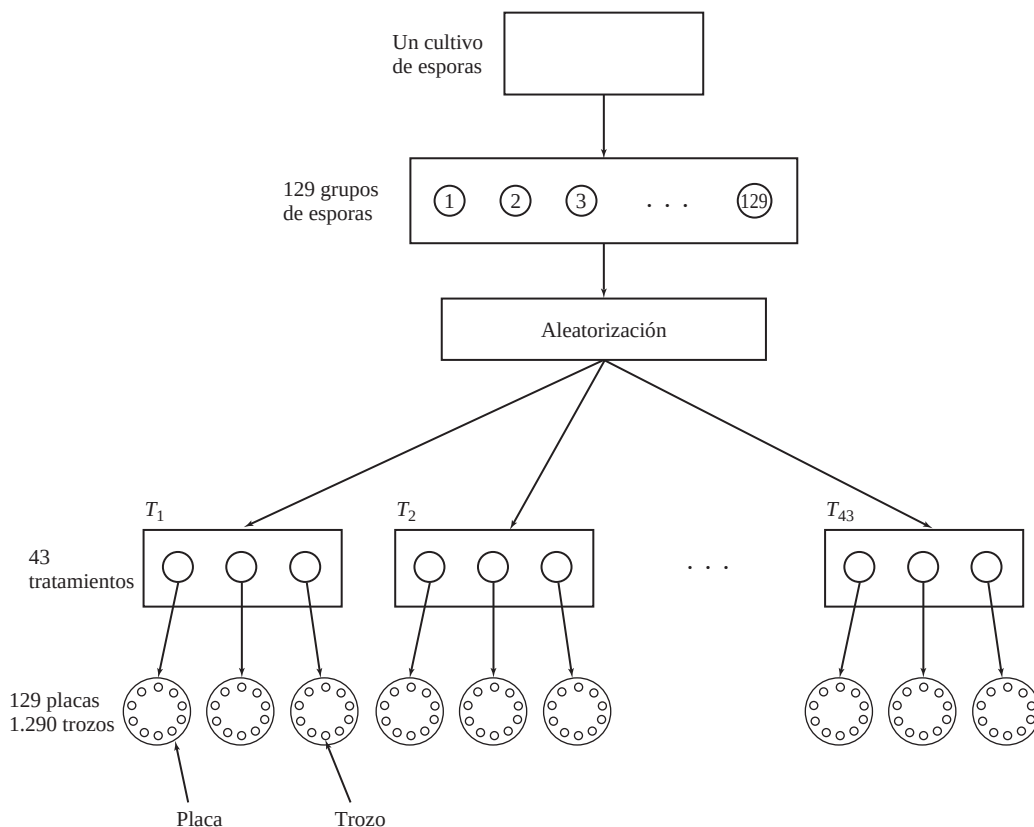


Figura 6.5.2 Diseño del experimento de germinación de esporas

Para tomar conciencia de los aspectos que surgen en este diseño, observemos algunos de los datos. La Tabla 6.5.1 muestra el porcentaje de esporas que había germinado en cada trozo del tratamiento 1.

La Tabla 6.5.1 muestra que hay una variabilidad considerable tanto *dentro de* cada placa de Petri como *entre* las placas. La variabilidad dentro de las placas refleja la variación local en el porcentaje de germinación, debida quizá en gran parte a diferencias entre las propias esporas (algunas de las esporas eran más maduras que otras). La variabilidad entre las placas es incluso mayor, debido que incluye no solo la variación local, sino también variación a mayor escala, como la variabilidad entre los grupos iniciales de las esporas y variaciones en la temperatura y humedad relativa dentro de las cámaras de almacenamiento.

Tabla 6.5.1 Porcentaje de germinación bajo el tratamiento 1			
	Placa I	Placa II	Placa III
	49	66	49
	58	84	60
	48	83	54
	69	69	72
	45	72	57
	43	85	70
	60	59	65
	44	60	68
	44	75	66
	68	68	60
Media	52,8	72,1	62,1
DT	10,1	9,5	7,4

Consideremos ahora el problema de comparar el tratamiento 1 con los otros tratamientos. ¿Sería legítimo tomar el punto de vista de que tenemos 30 observaciones en cada tratamiento? Para enfocar la pregunta, consideremos el cálculo del error típico de la media del tratamiento 1. La media y DT de las 30 observaciones son:

$$\text{Media} = 62,33$$

$$\text{ET} = 11,88$$

¿Sería legítimo calcular el ET de la media así?:

$$\text{ET}_{\bar{y}} = \frac{s}{\sqrt{n}} = \frac{11,88}{\sqrt{30}} = 2,2$$

Como podemos sospechar, **esto no es legítimo**. Hay una estructura jerárquica en los datos y, por tanto, no podemos aplicar tan inocentemente la fórmula del ET. Una forma aceptable de calcular el ET es considerar la media de cada placa como una observación. De esta forma obtenemos lo siguiente*:

Observaciones: 52,8, 72,1, 62,1

$$n = 3$$

$$\text{Media} = 62,33$$

$$\text{ET} = 9,65$$

$$\text{ET}_{\bar{y}} = \frac{s}{\sqrt{n}} = \frac{9,65}{\sqrt{3}} = 5,6$$

Nótese que el análisis incorrecto da como resultado la misma media (62,33) que este análisis, pero un valor de ET inapropiadamente pequeño (2,2 en vez de 5,6). Si fuéramos a comparar varios tratamientos, tendería a aparecer el mismo patrón: el análisis incorrecto tendería a producir valores de ET que serían (de forma individual o agregada) demasiado pequeños, lo que podría causar que «sobreinterpretáramos» los datos, en el sentido de sugerir que hay una evidencia significativa de diferencias en los tratamientos donde no existe ninguna.

* Una forma alternativa de arreglar los datos de los 10 trozos de una placa sería combinar los datos globales de esporas germinadas y no germinadas de la placa completa y expresarlos como porcentaje global de germinación.

Debemos resaltar que incluso aunque el análisis correcto requiere combinar las medidas de los 10 trozos de una placa en una única observación de esa placa, no ha sido en vano el esfuerzo de medir 10 trozos por placa en vez de, por ejemplo, solo un trozo por placa. La media de 10 trozos es un estimador mucho mejor del promedio de la placa completa que una sola medida de la placa. La mejora de precisión al medir 10 trozos se refleja en una DT entre placas menor. Por ejemplo, para el tratamiento 1 la DT fue de 9,65; si se hubieran medido menos trozos por placa, esta DT hubiera sido probablemente mayor. ■

El problema ilustrado en el Ejemplo 6.5.3 ha atrapado a muchos investigadores confiados. Cuando aparecen estructuras jerárquicas a partir de muchas medidas repetidas del mismo organismo individual (como en el Ejemplo 6.5.2) son relativamente fáciles de reconocer. Pero la estructura jerárquica del Ejemplo 6.5.3 tiene un origen diferente. Es debida al hecho de que la unidad observacional es un trozo individual, pero los trozos individuales no se asignaron aleatoriamente a los grupos de tratamiento. En vez de eso, la unidad que se asignó aleatoriamente al tratamiento es un grupo de esporas, que más tarde se siembra en una placa de Petri, que después dé lugar a 10 trozos. En el lenguaje de diseño de experimentos, los trozos están **anidados** dentro de las placas de Petri. *Siempre que las unidades observacionales estén anidadas dentro de las unidades que se asignaron aleatoriamente a los tratamientos, puede aparecer potencialmente una estructura jerárquica de los datos.* Nótese que la dificultad es solamente «potencial». En algunos casos un análisis no jerárquico puede ser aceptable. Por ejemplo, si la experiencia hubiera demostrado que las diferencias entre las placas de Petri fueran despreciables, entonces podríamos ignorar la estructura jerárquica en el análisis de los datos. La decisión puede ser difícil y puede requerir la ayuda de expertos en estadística.

El asunto de las estructuras jerárquicas de datos tiene importantes implicaciones en el diseño del experimento, así como en el análisis. El tamaño de la muestra (n) se debe identificar apropiadamente para determinar si el experimento incluye suficientes réplicas. Como un ejemplo sencillo, suponga que se propone realizar un experimento de germinación de esporas como el del Ejemplo 6.5.3, pero con solo *una* placa por tratamiento, en vez de tres. Para ver el efecto de esta propuesta, supongamos que el experimento planteado va a incluir tres tratamientos, con una placa por tratamiento. Con este diseño, ¿podríamos distinguir diferencias entre tratamientos de diferencias inherentes entre las placas? No. Las diferencias entre tratamientos y las diferencias entre las placas están enredadas mutuamente, o confundidas. Podemos visualizar esta situación fácilmente si observamos los datos de la Tabla 6.5.1 y suponemos que esos datos provinieron del experimento propuesto. Es decir, suponemos que las placas I, II y III han recibido diferentes tratamientos, y que no tenemos otros datos. Sería difícil extraer información con algún sentido sobre las diferencias entre tratamientos, a menos que supiéramos de forma *cierta* que la variación entre las placas es despreciable.

En la Sección 6.4 vimos cómo utilizar una estimación preliminar de la DT para determinar el tamaño de la muestra (n) requerido para obtener un cierto grado de precisión, expresada por el ET. Estas ideas se trasladan a los experimentos en los que aparecen estructuras de datos jerárquicas. Por ejemplo, supongamos que una botánica está planeando un experimento de germinación de esporas como el del Ejemplo 6.5.3. Si ya hubiera decidido usar 10 trozos por placa, el problema restante sería decidir el número de placas por tratamiento. Esta cuestión se podría plantear como en la Sección 6.4, considerando la placa como la unidad experimental, y utilizando una estimación preliminar de la DT entre placas (que en el Ejemplo 6.5.3 varía 9,65). Sin embargo, si deseara escoger valores óptimos *tanto* para el número de trozos por placa *como* para el número de placas por tratamiento, debería consultar a un estadístico.

Condiciones para la validez de un intervalo de confianza para μ

Un intervalo de confianza para μ proporciona una interpretación cuantitativa definitiva del $ET_{\bar{y}}$. Nótese que los datos deben ser una muestra aleatoria de la población de interés. Si hay un sesgo en el proceso de muestreo, entonces los conceptos de muestreo de distribuciones en los que se basa el método del intervalo de confianza no se cumplen. Conocer la media de una muestra sesgada no proporciona información sobre la media poblacional μ . La validez del método de la t de Student para construir intervalos de confianza depende también de la forma de la distribución poblacional de la variable observada Y . Si Y siguiera una distribución normal en la población, entonces el método de la t de Student es **exactamente válido**: es decir, la probabilidad de que el intervalo de confianza contenga a μ es exactamente igual al nivel de confianza (por ejemplo, 95 %). De la misma forma, esta interpretación es aproximadamente válida si la distribución poblacional es aproximadamente normal. Incluso si la distribución poblacional no es normal, el intervalo de confianza basado en la t de Student es aproximadamente válido *si* el tamaño de la muestra es grande. Este hecho se puede utilizar a

menudo para justificar el uso del intervalo de confianza incluso en situaciones donde no se puede asumir que la distribución poblacional sea aproximadamente normal.

Desde un punto de vista práctico, la pregunta importante es: ¿Qué tamaño debe tener la muestra para que el intervalo de confianza sea aproximadamente válido? De forma no sorprendente, la respuesta a esta pregunta depende del *grado* de no normalidad de la distribución poblacional: si la población es solo moderadamente no normal, entonces no es necesario que n sea muy grande. La Tabla 6.5.2 muestra la probabilidad real de que un intervalo de confianza basado en la t de Student contenga a μ para muestras de tres poblaciones diferentes²⁷. Las formas de las distribuciones poblacionales se muestran en la Figura 6.5.3.

Tabla 6.5.2 Probabilidad real de que los intervalos de confianza contengan a la media poblacional							
(a) Intervalo de confianza del 95 %							
	Tamaño de la muestra						
	2	4	8	16	32	64	Muy grande
Población 1	0,95	0,95	0,95	0,95	0,95	0,95	0,95
Población 2	0,94	0,93	0,94	0,94	0,95	0,95	0,95
Población 3	0,87	0,53	0,57	0,80	0,88	0,92	0,95
(b) Intervalo de confianza del 99 %							
	Tamaño de la muestra						
	2	4	8	16	32	64	Muy grande
Población 1	0,99	0,99	0,99	0,99	0,99	0,99	0,99
Población 2	0,99	0,98	0,98	0,98	0,99	0,99	0,99
Población 3	0,97	0,82	0,60	0,81	0,93	0,96	0,99

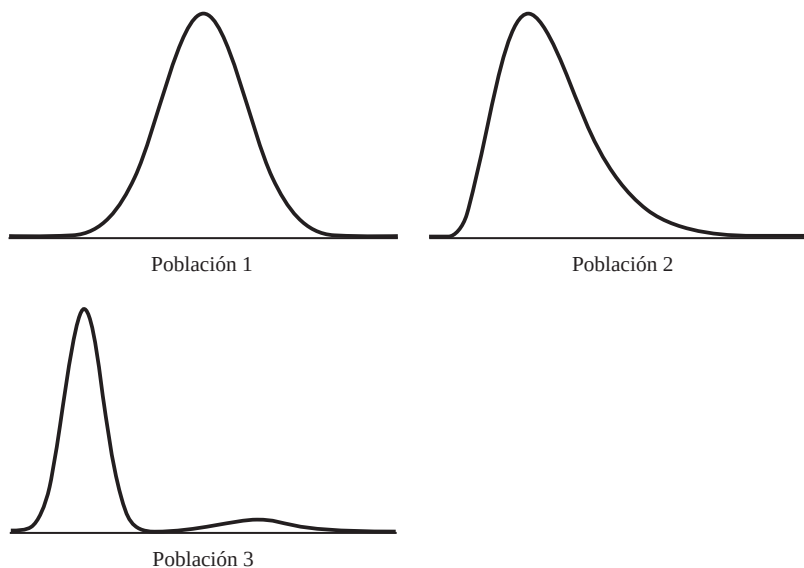


Figura 6.5.3 Tres distribuciones poblacionales: (1) normal, (2) ligeramente sesgada hacia la derecha, (3) altamente sesgada hacia la derecha

La población 1 es una población normal, la población 2 está moderadamente sesgada y la población 3 está extremadamente sesgada, una distribución con «forma de L». [Las poblaciones 2 y 3 se explicaron en la Sección 5.3 (opcional)].

En el caso de la población 2, el método es aproximadamente válido incluso para muestras bastante pequeñas. Para la población 3 la aproximación es muy pobre para muestras pequeñas y solamente es aceptable para muestras tan grandes como $n = 64$. En cierto sentido, la población 3 es un «caso peor». Se podría argumentar que la media no es un estimador significativo de la población 3, debido a su extraña forma.

Resumen de condiciones

En resumen, el método de la t de Student para construir un intervalo de confianza para μ es apropiado si se mantienen las siguientes condiciones.

1. Condiciones sobre el diseño del estudio

- (a) Debe ser razonable considerar los datos como una muestra aleatoria de una población grande.
- (b) Las observaciones de la muestra deben ser independientes entre sí.

2. Condiciones sobre la forma de la distribución poblacional

- (a) Si n es pequeño, la distribución poblacional debe ser aproximadamente normal.
- (b) Si n es grande, la distribución poblacional no necesita ser aproximadamente normal.

El requisito de que los datos sean una muestra aleatoria es la condición más importante.

El significado de «grande» en la condición 2(b) depende (como se muestra en el Ejemplo 6.5.3) del grado de no normalidad de la población. En muchas situaciones prácticas, un tamaño de muestra moderado (por ejemplo, $n = 20$ o 30) es suficientemente grande.

Verificación de condiciones

En la práctica, las «condiciones» precedentes son a menudo «suposiciones» en vez de hechos conocidos. Sin embargo, siempre es importante comprobar si las condiciones son razonables en un caso dado.

Para determinar si el modelo del muestreo aleatorio es aplicable a un estudio en particular, se debe examinar el diseño del estudio, con particular atención a posibles sesgos en la selección del material experimental y a una posible no independencia de las observaciones debida a estructuras jerárquicas de datos.

Con respecto a si la distribución poblacional es aproximadamente normal, la información sobre este punto puede estar disponible en la experiencia previa con datos similares. Si la única fuente de información son los datos que se tienen en ese momento, entonces la normalidad se puede comprobar de forma aproximada realizando el histograma y una gráfica de probabilidad normal de los datos. Desafortunadamente, para un tamaño de la muestra pequeño o moderado, esta comprobación es bastante burda. Por ejemplo, si volvemos a la Figura 5.2.7, podemos ver que incluso muestras de tamaño 25 de una población normal no parecen frecuentemente ser particularmente normales*. Por supuesto, si la muestra es grande, entonces el histograma de dicha muestra nos da una información buena sobre la forma de la población. Sin embargo, si n es grande, el requisito de normalidad es menos importante en cualquier caso.

De todas formas, una comprobación burda es mejor que ninguna, y *cada* análisis de datos debería empezar con una inspección gráfica de dichos datos, con especial atención a cualquier observación que esté muy lejos del centro de la distribución.

Algunas veces un histograma o una gráfica de probabilidad normal de los datos indican que dichos datos no provienen de una población normal. Si el tamaño de la muestra es pequeño, entonces el método de la t de Student no proporcionará resultados válidos. Sin embargo, puede ser posible transformar los datos para obtener una normalidad aproximada y después analizar dichos datos en la escala transformada.

* Podríamos ayudar a nuestra evaluación gráfica de la normalidad utilizando un método más objetivo como la prueba de Shapiro-Wilk de la Sección 4.4.

Ejemplo 6.5.4 Rendimiento de sedimentos

El rendimiento de sedimentos, que es una medida de la cantidad de sedimentos suspendidos en el agua, es una medida de la calidad del agua de un río. La distribución del rendimiento de sedimentos a menudo tiene una distribución sesgada. Sin embargo, tomando el logaritmo de cada observación se puede producir una distribución que sigue una curva normal bastante bien. La Figura 6.5.4 muestra las gráficas de probabilidad normal de los rendimientos de sedimentos de muestras de agua del Black River en el noreste de Ohio para $n = 9$ días (a) en mg/l y (b) en escala logarítmica (es decir, $\ln(\text{mg/l})$)²⁸.

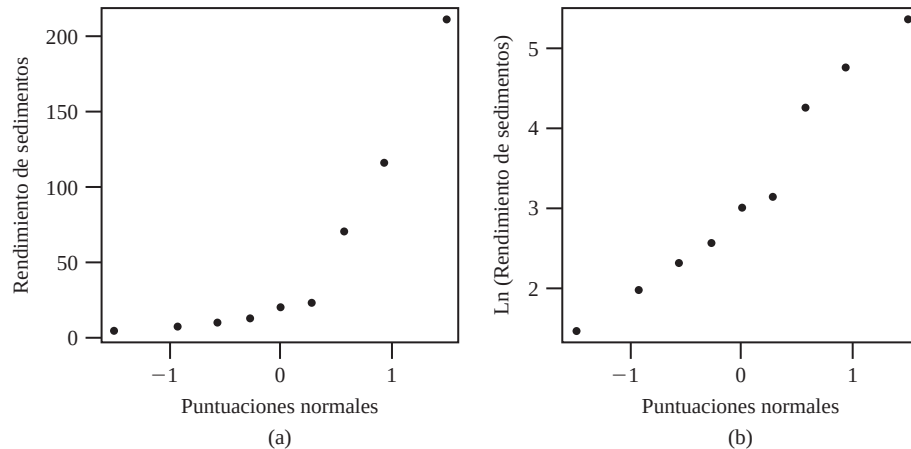


Figura 6.5.4 Gráficas de probabilidad normal de los rendimientos de sedimentos de agua del Black River en nueve días (a) en mg/l y (b) tras tomar el logaritmo natural de cada observación*

Los logaritmos naturales de los rendimientos de sedimentos tienen una media de $\bar{y} = 3,21$ y una desviación típica de $s = 1,33$. Por tanto, el error típico de la media es $\frac{1,33}{\sqrt{9}} = 0,44$. El multiplicador t para un intervalo de confianza del 95 % es $t_{8,0,025} = 2,306$. Un intervalo de confianza del 95 % para μ es

$$3,21 \pm 2,306(0,44)$$

o, aproximadamente,

$$3,21 \pm 1,01$$

o

$$(2,20, 4,22)$$

Por tanto, tenemos una confianza del 95 % de que la media del logaritmo natural del rendimiento de sedimentos del Black River esté entre 2,20 y 4,22[†]. ■

* La prueba de normalidad de Shapiro-Wilk (de la Sección 4.4) para los datos da un P valor de 0,0039, proporcionando así una fuerte evidencia de la no normalidad de los datos sin transformar. Por el contrario, para los datos transformados con el logaritmo natural, el P valor es 0,6551, mostrando así que no hay evidencia significativa de no normalidad. Nótese que podríamos haber utilizado también el logaritmo en base 10 para normalizar los datos.

† Nótese que hemos construido un intervalo de confianza para la media del logaritmo de la población del rendimiento de sedimentos. Debido a que la transformación logarítmica no es lineal, la media del logaritmo no es el logaritmo de la media, por lo que aplicar la transformación inversa a los extremos del intervalo de confianza no lo convertirá apropiadamente en un intervalo de confianza para la media poblacional en la escala original de mg/l. Sin embargo, podemos obtener un intervalo de confianza aproximado tomando $\exp(2,2 + 1,33^2/2)$ y $\exp(4,22 + 1,33^2/2)$. [Esto se basa en el hecho de que la media de una distribución log normal (que tiene forma de campana después de tomar logaritmos) es $\exp(\mu + \sigma^2/2)$].

Ejercicios 6.5.1-6.5.8

6.5.1 La transaminasa glutámico-oxalacética sérica (SGOT, de sus siglas en inglés) es una enzima que muestra una actividad elevada cuando el músculo cardíaco está dañado. En un estudio de 31 pacientes que recibieron cirugía cardíaca, se midieron sus niveles en suero de SGOT 18 horas después de la cirugía²⁹. La media fue de 49,3 U/l y la desviación típica fue de 68,3 U/l. Si consideramos las 31 observaciones como una muestra de una población, ¿qué característica de los datos haría aparecer dudas sobre la distribución normal de dicha población?

6.5.2 Un árbol dendrítico es una estructura ramificada que surge del cuerpo de una célula nerviosa. En un estudio sobre el desarrollo del cerebro, unos investigadores examinaron tejidos cerebrales de siete cerdos guinea adultos. Los investigadores seleccionaron aleatoriamente células nerviosas de una cierta región del cerebro y contaron el número de segmentos de ramas dendríticas que surgían de cada célula seleccionada. Se seleccionaron un total de 36 células, y los números resultantes fueron las siguientes³⁰:

38	42	25	35	35	33	48	53	17
24	26	26	47	28	24	35	38	26
38	29	49	26	41	26	35	38	44
25	45	28	31	46	32	39	59	53

La media de estos números es 35,67 y la desviación típica es 9,99.

Suponga que se desea construir un intervalo de confianza del 95 % para la media poblacional. Podríamos calcular el error típico como

$$ET_{\bar{y}} = \frac{9,99}{\sqrt{36}} = 1,67$$

y obtener el intervalo de confianza como

$$35,67 \pm (2,042)(1,67)$$

o

$$32,3 < \mu < 39,1$$

- (a) ¿Con qué base se podría criticar el análisis anterior? (Sugerencia: ¿son las observaciones independientes?).
- (b) Utilizando las clases [15,20), [20, 25) y así sucesivamente, construya un histograma de los datos. ¿Da fundamento la forma de la distribución a la crítica que realizó en el apartado (a)? Si es así, explique cómo.

6.5.3 En un experimento para estudiar la regulación de la secreción de insulina se tomaron muestras de sangre de siete perros antes y después de la estimulación eléctrica del nervio vago. Los valores siguientes muestran, para cada animal, el incremento (después menos antes) en la concentración de insulina inmunorreactiva ($\mu\text{U/ml}$) en el plasma de la vena pancreática³¹.

30	100	60	30	130	1.060	30

Para estos datos, el método de la t de Student produce el siguiente intervalo de confianza del 95 % para la media poblacional:

$$-1,45 < \mu < 556$$

¿Es apropiado el método de la t de Student en este caso? ¿Por qué o por qué no?

6.5.4 En un estudio de las relaciones entre parásito y huésped, se expusieron 242 larvas de la palomilla *Ephestia* a parasitación por la mosca *Ichneumon*. La tabla siguiente muestra el número de huevos de *Ichneumon* encontrados en cada una de las larvas de *Ephestia*³².

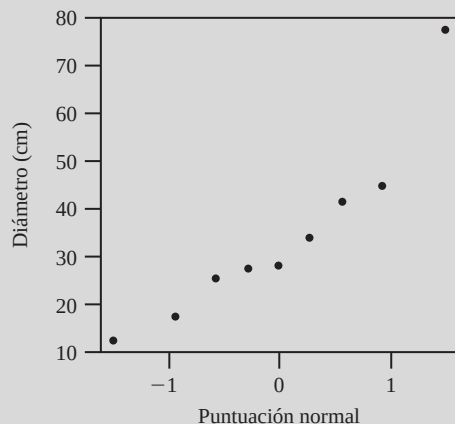
N.º de huevos (y)	N.º de larvas
0	21
1	77
2	52
3	41
4	23
5	13
6	9
7	1
8	2
9	0
10	2
11	0
12	0
13	0
14	0
15	1
Total	242

Para estos datos, $\bar{y} = 2,368$ y $s = 1,950$. El método de la t de Student proporciona el siguiente intervalo de confianza del 95 % para μ , la media poblacional del número de huevos por larva:

$$2,12 < \mu < 2,61$$

- (a) ¿Parece razonable suponer que la distribución poblacional de Y es aproximadamente normal? Explique su respuesta.
- (b) En vista de su respuesta al apartado (a), ¿sobre qué base podría defender la aplicación del método de la t de Student a estos datos?

6.5.5 La siguiente gráfica de probabilidad normal muestra la distribución de los diámetros, en cm, de siete árboles Si-cómos Americanos³³.



La gráfica de probabilidad normal no es lineal, lo que sugiere que es necesaria una transformación de los datos antes de que se pueda construir un intervalo de confianza utilizando el método de la t de Student. Los datos son:

12,4 44,8 28,2 77,6 34 17,5 41,5 25,5 27,5

- Tome la raíz cuadrada de cada observación y construya un intervalo de confianza del 90 % para la media.
- Interprete el intervalo de confianza del apartado (a). Es decir, explique lo que indica el intervalo de confianza sobre la raíz cuadrada de los diámetros de estos árboles.

6.5.6 Se compararon cuatro tratamientos para ver su efecto sobre el crecimiento de células de espinaca en matraces de cultivo de células. El investigador asignó aleatoriamente

dos matraces a cada tratamiento. Después de un cierto tiempo de tratamiento, seleccionó aleatoriamente tres partes alícuotas (de 1 cc cada una) de cada matraz y midió la densidad de células de cada una. Por tanto, tenía seis medidas de densidad de células en cada tratamiento. Para obtener el error típico de la media de un tratamiento, el investigador calculó la desviación típica de las seis medidas y dividió por $\sqrt{6}$. ¿Sobre qué base se podría plantear una objeción este método de calcular el ET?

6.5.7 En un experimento sobre variedades de soja se hicieron crecer en invernadero plantas de soja plantadas individualmente, teniendo 10 plantas cada variedad utilizada en el experimento. Se seleccionaron de forma aleatoria cinco semillas de la cosecha de cada planta y se analizó individualmente su porcentaje de aceite. Esto proporciona un total de 50 medidas para cada variedad. Para calcular el error típico de la media de una variedad, el experimentador calculó la desviación típica de las 50 observaciones y dividió por $\sqrt{50}$. ¿Por qué sería de dudosa validez este cálculo?

6.5.8 En un proyecto sobre la extinción de plantas, una población local completa (en peligro de extinción) de 255 arbustos Congdon se trasplantó a una nueva localización³⁴. Un año después del trasplante se seleccionaron aleatoriamente 30 de las 255 plantas y se midió el diámetro de la unión *caudix* de la raíz (la parte alta de la raíz justo por debajo de la superficie del suelo). Si la población de plantas bajo consideración está formada solo por las 255 plantas locales, explique por qué no sería apropiado utilizar el método de la t de Student para construir un intervalo de confianza para μ , la media poblacional de los diámetros de las uniones de la raíz.

6.6 Comparación de dos medias

En las secciones anteriores hemos considerado el análisis de una única muestra de datos cuantitativos. Sin embargo, en la práctica, muchas investigaciones científicas requieren la comparación de dos o más muestras de diferentes poblaciones. Cuando la variable observada es cuantitativa, la comparación de dos muestras puede incluir diversos aspectos, particularmente (1) comparación de medias, (2) comparación de desviaciones típicas y (3) comparación de formas. En esta sección, y en realidad en todo el libro, el objetivo principal será la comparación de medias y otras comparaciones relacionadas con cambios. Comenzaremos presentando el planteamiento del intervalo de confianza para comparar medias, que es una extensión natural del material de la Sección 6.3. En el Capítulo 7 consideraremos un planteamiento conocido como contraste de hipótesis.

Notación

La Figura 6.6.1 presenta nuestra notación para la comparación de dos muestras. La notación es exactamente paralela a nuestra notación anterior, pero con un subíndice (1 o 2) para diferenciar entre las dos muestras. Las dos «poblaciones» pueden ser poblaciones que aparecen de forma natural (como en el Ejemplo 6.1.1) o pueden ser poblaciones conceptuales definidas por ciertas condiciones experimentales (como en el Ejemplo 6.3.4). En cada caso, los datos de cada muestra se ven como una muestra aleatoria de la correspondiente población. Comenzaremos por describir, en la sección siguiente, algunos cálculos simples que se utilizan tanto en los intervalos de confianza como en el contraste de hipótesis.

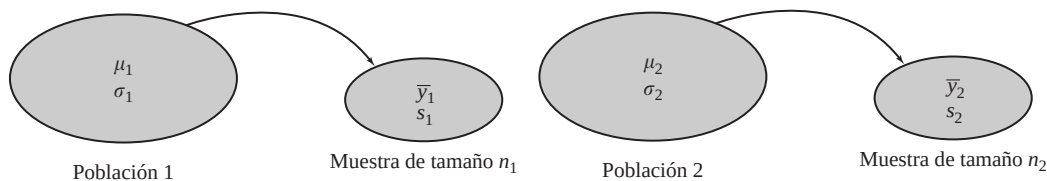


Figura 6.6.1 Notación para comparación de dos muestras

Error típico de $(\bar{Y}_1 - \bar{Y}_2)$

En esta sección presentaremos una cantidad fundamental para comparar dos muestras: el error típico de la diferencia entre las medias muestrales.

Ideas básicas

En este capítulo hemos visto que la precisión de una media muestral $\bar{Y}$ se puede expresar mediante su error típico, que es igual a

$$ET_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

Para comparar dos medias muestrales, es natural considerar la diferencia entre ellas:

$$\bar{Y}_1 - \bar{Y}_2$$

que es un estimador de la cantidad $(\mu_1 - \mu_2)$. Para caracterizar el error de muestreo de la estimación, necesitamos analizar el error típico de la diferencia $(\bar{Y}_1 - \bar{Y}_2)$. Ilustraremos esta idea con un ejemplo.

Ejemplo 6.6.1 Capacidad vital

La capacidad vital es una medida de la cantidad de aire que se puede exhalar después de hacer una inspiración profunda. Se podría esperar que los músicos que tocan instrumentos de metal tuvieran capacidades vitales más grandes, en promedio, que la de otras personas de la misma edad, sexo y altura. En un estudio se compararon las capacidades vitales de ocho músicos que tocaban instrumentos de metal con las capacidades vitales de siete sujetos de control. La Tabla 6.6.1 muestra los datos³⁵.

La diferencia entre las medias muestrales es

$$\bar{y}_1 - \bar{y}_2 = 4,83 - 4,74 = 0,09$$

Sabemos que tanto $\bar{y}_1$ como $\bar{y}_2$ están sujetas a error de muestreo y, en consecuencia, la diferencia (0,09) está también sujeta al error de muestreo. El error típico de $(\bar{Y}_1 - \bar{Y}_2)$ nos indica cuánta precisión se puede asignar a esta diferencia entre $\bar{Y}_1$ e $\bar{Y}_2$. ■

Definición El **error típico de $\bar{Y}_1 - \bar{Y}_2$** se define como

$$ET_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

La forma alternativa siguiente de la fórmula muestra cómo el ET de la diferencia se relaciona con cada uno de los ET de las medias:

$$ET_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{ET_1^2 + ET_2^2}$$

Tabla 6.6.1 Capacidades vitales (litros)		
	Músicos de instrumentos de metal	Control
	4,7	4,2
	4,6	4,7
	4,3	5,1
	4,5	4,7
	5,5	5,0
	4,9	
	5,3	
n	7	5
$\bar{y}$	4,83	4,74
s	0,435	0,351

siendo

$$ET_1 = ET_{\bar{Y}_1} = \frac{s_1}{\sqrt{n_1}}$$

$$ET_2 = ET_{\bar{Y}_2} = \frac{s_2}{\sqrt{n_2}}$$

Nótese que esta versión de la fórmula muestra que «los ET se suman como Pitágoras». Cuando tenemos dos muestras independientes, tomamos el ET de cada media, los elevamos al cuadrado, lo sumamos y después tomamos la raíz cuadrada de la suma. La Figura 6.6.2 ilustra esta idea.

Puede parecer extraño que al calcular el ET de una diferencia *sumemos* en vez de restar en la fórmula $ET_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{ET_1^2 + ET_2^2}$. Sin embargo, como se comentó en la Sección 3.5, la variabilidad de la diferencia depende de la variabilidad de cada parte. Tanto si sumamos $\bar{Y}_2$ con $\bar{Y}_1$ como si restamos $\bar{Y}_2$ de $\bar{Y}_1$, el «ruido» asociado con $\bar{Y}_2$ (es decir, ET_2) se suma a la incertidumbre global. Cuanto mayor sea la variabilidad de $\bar{Y}_2$, mayor será la variabilidad de $(\bar{Y}_1 - \bar{Y}_2)$. La fórmula $ET_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{ET_1^2 + ET_2^2}$ da cuenta de esta variabilidad.

Ilustraremos las fórmulas en el siguiente ejemplo.

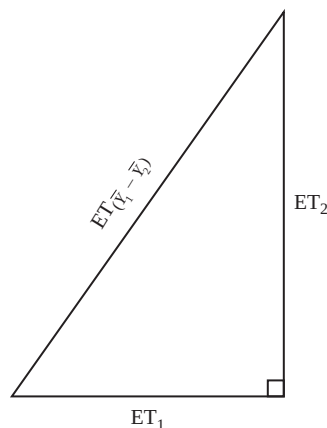


Figura 6.6.2 ET de una diferencia

Ejemplo 6.6.2 Capacidad vital

Para los datos de capacidad vital, los cálculos preliminares producen los resultados que se muestran en la Tabla 6.6.2.

El ET de $(\bar{Y}_1 - \bar{Y}_2)$ es

$$ET_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{0,1892}{7} + \frac{0,1232}{5}} = 0,227 \approx 0,23$$

Adviértase que

$$0,227 = \sqrt{(0,164)^2 + (0,157)^2}$$

Nótese que el ET de la diferencia es mayor que cada uno de los ET individuales, pero menor que su suma. ■

Tabla 6.6.2		
	Músicos de instrumentos de metal	Control
s^2	0,1892	0,1232
n	7	5
ET	0,164	0,157

Ejemplo 6.6.3 Tonsilectomía

Se realizó un experimento para comparar la cirugía convencional con un procedimiento nuevo denominado tonsilectomía intracapsular asistida por coblación, en niños a los que había que extirpar su lengua. Una medida clave tomada durante el estudio fue la puntuación de dolor que indicó cada niño, en una escala de 0-10, cuatro días después de la cirugía. La Tabla 6.6.3 muestra las medias y las desviaciones típicas de las puntuaciones de dolor de los dos grupos³⁶.

Tabla 6.6.3 Puntuación de dolor		
	Tipo de cirugía	
	Convencional	Coblación
Media	4,3	1,9
DT	2,8	1,8
n	49	52

Los datos de la Tabla 6.6.3 muestran que la desviación típica de las puntuaciones de dolor en 49 niños que recibieron cirugía convencional fue de 2,8. Por tanto, el ET de la media convencional es $\frac{2,8}{\sqrt{49}} = 0,40$. Para los 52 niños del grupo de la coblación, la DT fue 1,8, con lo que el ET es $\frac{1,8}{\sqrt{52}} = 0,2496$. El ET de la diferencia de las dos medias es $\sqrt{0,40^2 + 0,25^2} = 0,4717 \approx 0,47$. ■

El error típico agregado (opcional)

El error típico anterior se conoce como error típico «desagrupado». Muchos paquetes de *software* estadístico permiten al usuario especificar el uso de lo que se conoce como el error típico «agrupado», que presentaremos brevemente.

Recuérdese que el cuadrado de la desviación típica, s , es la varianza muestral, s^2 , definida como

$$s^2 = \frac{\sum (\bar{y}_i - \bar{y})^2}{n - 1}$$

La varianza agrupada es un promedio ponderado de s_1^2 , la varianza de la primera muestra, y s_2^2 , la varianza de la segunda muestra, con pesos iguales a los grados de libertad de cada muestra, $n_i - 1$:

$$s_{\text{agrupada}}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

El error típico agrupado se define como

$$ET_{\text{agregado}} = \sqrt{s_{\text{agregada}}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Presentamos un ejemplo ilustrativo.

Ejemplo 6.6.4 Capacidad vital

Para los datos de capacidad vital obtuvimos que $s_1^2 = 0,1892$ y $s_2^2 = 0,1232$. La varianza agrupada es

$$s_{\text{agrupada}}^2 = \frac{(7 - 1)0,1892 + (5 - 1)0,1232}{(7 + 5 - 2)} = 0,1628$$

y el ET agrupado es

$$ET_{\text{agregado}} = \sqrt{0,1628 \left(\frac{1}{7} + \frac{1}{5} \right)} = 0,236$$

Recuérdese del Ejemplo 6.6.2 que el ET desagrupado para los mismos datos era 0,227. ■

Si los tamaños de las muestras son iguales ($n_1 = n_2$) o si las desviaciones típicas son iguales ($s_1 = s_2$), entonces el método agrupado y el desagrupado producirán el mismo resultado para $ET_{(\bar{Y}_1 - \bar{Y}_2)}$. Las dos respuestas no se diferenciarán sustancialmente a menos que tanto los tamaños de las muestras como las DT muestrales sean bastante distintas.

Para mostrar la analogía entre las dos fórmulas del ET, podemos expresarlas como sigue:

$$ET_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

y

$$ET_{\text{agrupado}} = \sqrt{\frac{s_{\text{agrupada}}^2}{n_1} + \frac{s_{\text{agrupada}}^2}{n_2}}$$

En el método agrupado, las varianzas separadas (s_1^2 y s_2^2), se sustituyen por una única varianza s_{agrupada}^2 , que se calcula con ambas muestras.

Tanto el ET desagrupado como el agrupado tienen el mismo propósito: estimar la desviación típica de la distribución muestral de $(\bar{Y}_1 - \bar{Y}_2)$. De hecho, se puede demostrar que la desviación típica es

$$\sigma_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Nótese el parecido entre esta fórmula y la fórmula de $ET_{(\bar{Y}_1 - \bar{Y}_2)}$.

Al analizar datos cuando los tamaños de las muestras no son iguales ($n_1 \neq n_2$), es necesario decidir si utilizar el método agrupado o el desagrupado para el cálculo del error típico. La elección depende si se desea asumir que las DT poblacionales (σ_1 y σ_2) son iguales. Se puede demostrar que si $\sigma_1 = \sigma_2$, entonces se debería usar el método agrupado, porque en este caso s_{agrupada} es el mejor estimador de la DT poblacional. Sin embargo, en este caso el método desagrupado

dará en general un ET que será bastante similar al proporcionado por el método agrupado. Si $\sigma_1 \neq \sigma_2$, entonces se debería utilizar el método desagrupado, porque en este caso s_{agrupada} no es un estimador de σ_1 ni de σ_2 , por lo que la agrupación no conseguiría nada. Debido a que los dos métodos son sustancialmente similares cuando $\sigma_1 = \sigma_2$ y el método agrupado no es válido cuando $\sigma_1 \neq \sigma_2$, muchos estadísticos prefieren el método desagrupado. Hay poco que ganar con la agrupación cuando dicha agrupación es apropiada y mucho que perder cuando la agrupación no es apropiada. Muchos paquetes *software* utilizan el método desagrupado por defecto; el usuario debe especificar el uso del método agrupado si desea agrupar las varianzas.

Ejercicios 6.6.1-6.6.9

6.6.1 Datos provenientes de dos muestras proporcionaron los siguientes resultados:

	Muestra 1	Muestra 2
n	6	12
$\bar{y}$	40	50
s	4,3	5,7

Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$.

6.6.2 Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$ para los siguientes datos:

	Muestra 1	Muestra 2
n	10	10
$\bar{y}$	125	217
s	44,2	28,7

6.6.3 Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$ para los siguientes datos:

	Muestra 1	Muestra 2
n	5	7
$\bar{y}$	44	47
s	6,5	8,4

6.6.4 Considere los datos del Ejercicio 6.6.3. Suponga que se doblan los tamaños de las muestras, pero las medias y las DT permanecen iguales, como sigue. Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$.

	Muestra 1	Muestra 2
n	10	14
$\bar{y}$	44	47
s	6,5	8,4

6.6.5 Datos provenientes de dos muestras proporcionaron los siguientes resultados:

	Muestra 1	Muestra 2
$\bar{y}$	96,2	87,3
ET	3,7	4,6

Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$.

6.6.6 Datos provenientes de dos muestras proporcionaron los siguientes resultados:

	Muestra 1	Muestra 2
n	22	21
$\bar{y}$	1,7	2,4
ET	0,5	0,7

Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$.

6.6.7 El Ejemplo 6.6.3 presentó medidas de dolor en niños a los que se había extirpado la lengua. Otra variable medida en ese experimento fue el número de dosis de Tylenol tomadas por los niños de los dos grupos. Esos datos son

	Tipo de cirugía	
	Convencional	Coblación
n	49	52
$\bar{y}$	3,0	2,3
DT	2,4	2,0

Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$.

6.6.8 En un experimento controlado se hicieron crecer dos variedades de lechuga durante 16 días. La tabla siguiente muestra el peso total en seco (en gramos) de las hojas de nueve plantas de la variedad «Salad Bowl» y de seis plantas de la variedad «Bibb»³⁷.

	Salad Bowl	Bibb
	3,06	1,31
	2,78	1,17
	2,87	1,72
	3,52	1,20

(Continúa)

	Salad bowl	Bibb
	3,81	1,55
	3,60	1,53
	3,30	
	2,77	
	3,62	
$\bar{y}$	3,259	1,413
s	0,400	0,220

Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$ para estos datos.

6.6.9 Algunos fabricantes de jabón venden jabones especiales «antibacterianos». Sin embargo, se podría esperar que el jabón ordinario también matara las bacterias. Para investigar esto, un investigador preparó una solución de jabón ordinario no antibacteriano y una solución de control de agua estéril. Las dos soluciones se pusieron en placas de Petri y se añadió la bacteria *E. coli*. Las placas se incubaron durante 24 horas y se contó el número de colonias de bacterias de cada placa³⁸. Los datos se presentan en la siguiente tabla.

	Control (grupo 1)	Jabón (grupo 2)
	30	76
	36	27
	66	16
	21	30
	63	26
	38	46
	35	6
	45	
n	8	7
$\bar{y}$	41,8	32,4
s	15,6	22,8
ET	5,5	8,6

Calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$ para estos datos.

6.7 Intervalo de confianza para $(\mu_1 - \mu_2)$

Una forma de comparar dos medias muestrales es construir un intervalo de confianza para la diferencia de las medias poblacionales [es decir, un intervalo de confianza para la cantidad $(\mu_1 - \mu_2)$]. Recuerdese que hemos visto en este mismo capítulo que un intervalo de confianza del 95 % para la media μ de una única población distribuida normalmente se construye como

$$\bar{y} \pm t_{0,025} \text{ET}_{\bar{y}}$$

Análogamente, un intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ se construye como

$$(\bar{y}_1 - \bar{y}_2) \pm t_{0,025} \text{ET}_{(\bar{y}_1 - \bar{y}_2)}$$

El valor crítico $t_{0,025}$ se determina a partir de la distribución t de Student utilizando un número de grados de libertad* dado por

$$gl = \frac{(\text{ET}_1^2 + \text{ET}_2^2)}{\text{ET}_1^4/(n_1 - 1) + \text{ET}_2^4/(n_2 - 1)} \quad (6.7.1)$$

siendo $\text{ET}_1 = s_1/\sqrt{n_1}$ y $\text{ET}_2 = s_2/\sqrt{n_2}$.

Por supuesto, calcular los grados de libertad utilizando la fórmula (6.7.1) es complicado y consume tiempo. La mayor parte del *software* de computador y algunas calculadoras gráficas utilizan la fórmula (6.7.1). Un método más simple para obtener un número aproximado de grados de libertad es utilizar el mínimo de $(n_1 - 1)$ y $(n_2 - 1)$. Esta opción produce un intervalo de confianza que es algo conservativo, en el sentido de que el verdadero nivel de confianza es un poco mayor que el 95 % cuando se utiliza $t_{0,025}$. Un tercer planteamiento es aproximar los grados de libertad como $n_1 + n_2 - 2$. Este planteamiento es algo relajado en el sentido de que el verdadero nivel de confianza es algo más pequeño que el 95 % cuando se utiliza $t_{0,025}$.

* Estrictamente hablando, la distribución necesaria para construir un intervalo de confianza en este caso depende de las desviaciones típicas poblacionales desconocidas σ_1 y σ_2 y no es una distribución t de Student. Sin embargo, una distribución t de Student con grados de libertad dados por la fórmula (6.7.1) es una aproximación muy buena. Esto se conoce algunas veces como método de Welch o método de Satterwaite.

Los intervalos con otros coeficientes de confianza se construyen de forma análoga. Por ejemplo, para obtener un intervalo de confianza del 90 % utilizaríamos $t_{0,05}$ en lugar de $t_{0,025}$.

El ejemplo siguiente ilustra la construcción del intervalo de confianza para $(\mu_1 - \mu_2)$.

Ejemplo 6.7.1 Plantas rápidas

La Planta Rápida de Wisconsin, *Brassica campestris*, tiene un ciclo de crecimiento muy rápido que la hace particularmente adecuada para el estudio de factores que afectan al crecimiento de las plantas. En un estudio de ese tipo se trataron siete plantas con la sustancia Ancymidol (ancy) y se compararon con ocho plantas de control a las que se le suministró agua ordinaria. Se midieron las alturas de las plantas, en cm, tras 14 días de crecimiento³⁹. Los datos se muestran en la Tabla 6.7.1.

La observación en paralelo de los diagramas de puntos y las gráficas de probabilidad normal (Figura 6.7.1) muestra que ambas distribuciones muestrales son razonablemente simétricas y con forma de campana. Además, podemos esperar de una distribución de alturas de plantas se ajuste bastante bien a una distribución normal, ya que las distribuciones de alturas a menudo siguen una curva normal. El diagrama de puntos muestra que la distribución con ancy tiene un pequeño desplazamiento con respecto a la distribución de control. La diferencia en las medias muestrales es $15,9 - 11,0 = 4,9$. El ET de la diferencia de las medias muestrales es

$$ET_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{4,8^2}{8} + \frac{4,7^2}{7}} = 2,46$$

Tabla 6.7.1 Altura tras 14 días de crecimiento de plantas de control y de ancy (cm)		
	Control (grupo 1)	Ancy (grupo 2)
	10,0	13,2
	13,2	19,5
	19,8	11,0
	19,3	5,8
	21,2	12,8
	13,9	7,1
	20,3	7,7
	9,6	
n	8	7
$\bar{y}$	15,9	11,0
s	4,8	4,7
ET	1,7	1,8

Utilizando la fórmula (6.7.1), se obtiene que los grados de libertad son 12,8:

$$gl = \frac{(1,7^2 + 1,8^2)^2}{1,7^4/7 + 1,8^4/6} = 12,8$$

Empleando un computador, podemos obtener que para un intervalo de confianza del 95 % el multiplicador t para 12,8 grados de libertad es $t_{12,8,0,025} = 2,164$. (Sin utilizar computador, podríamos redondear los grados de libertad a 12, en cuyo caso el multiplicador t es 2,179. Este cambio de 12,8 a 12 grados de libertad tiene poco efecto en el resultado final). La fórmula del intervalo de confianza da

$$(15,9 - 11,0) \pm (2,164)(2,46)$$

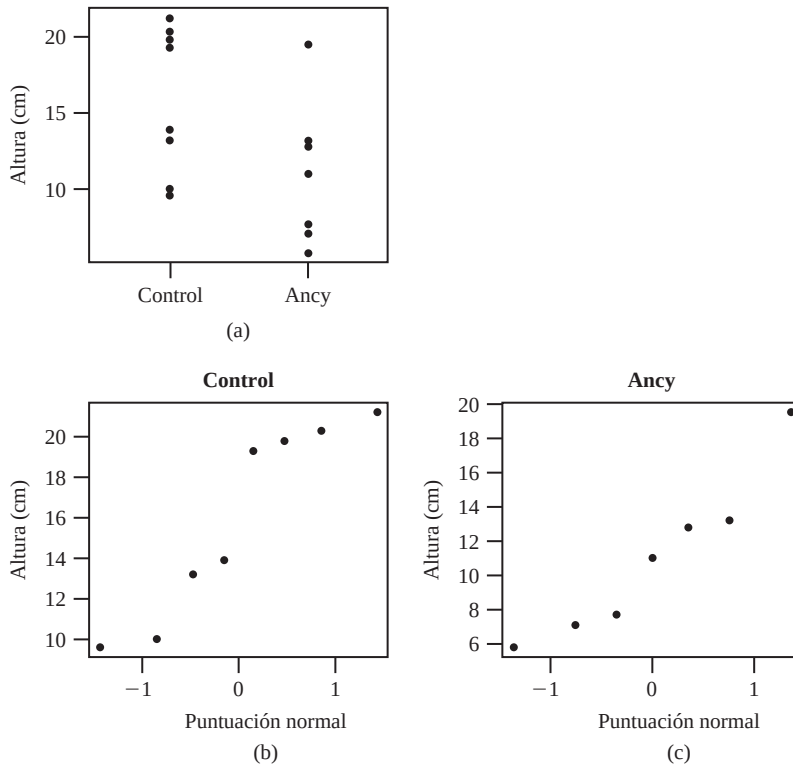


Figura 6.7.1 Presentación en paralelo del diagrama de puntos (a) y de las gráficas de probabilidad normal de las alturas de plantas rápidas de control (b) y de ancy (c)

o

$$4,9 \pm 5,32$$

El intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ es

$$(-0,42, 10,22)$$

Redondeando, tenemos

$$(-0,4, 10,2)$$

Por tanto, tenemos una confianza del 95 % de que la media poblacional de la altura tras 14 días de crecimiento de plantas rápidas cuando se utiliza agua (μ_1) esté entre 0,4 cm por debajo y 10,2 cm por encima que la media poblacional de la altura tras 14 días de crecimiento de plantas rápidas cuando se utiliza ancy (μ_2). ■

Ejemplo 6.7.2 Plantas rápidas

Dijimos que un método conservativo para construir un intervalo de confianza para diferencias de medias es utilizar el mínimo de $(n_1 - 1)$ y $(n_2 - 1)$. Para los datos dados en el Ejemplo 6.7.1, este método utilizaría 6 grados de libertad y un multiplicador t de 2,447. En este caso, el intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ es

$$(15,9 - 11,0) \pm (2,447)(2,46)$$

o

$$4,9 \pm 6,02$$

El intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ es

$$(-1,1, 10,9)$$

Este intervalo es un poco conservativo en el sentido de que es más amplio que el intervalo obtenido en el Ejemplo 6.7.1. ■

Ejemplo 6.7.3 Peso del tórax de mariposas

Los biólogos tienen la teoría de que las mariposas Monarch macho tienen, en promedio, un tórax más grande que las hembras. Una muestra de siete mariposas macho y ocho hembras produjo los datos de la Tabla 6.7.2, que se muestran en la Figura 6.7.2. (Estos datos vienen de otra parte del estudio descrito en el Ejemplo 6.1.1).

Para los datos de la Tabla 6.7.2, el ET para $(\bar{Y}_1 - \bar{Y}_2)$ es

$$ET_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{8,4^2}{7} + \frac{7,5^2}{8}} = 4,14$$

La fórmula (6.7.1) da el número de grados de libertad

$$gl = \frac{(3,2^2 + (2,7^2)^2)}{\frac{3,2^4}{6} + \frac{2,7^4}{7}} = 12,3$$

Para un intervalo de confianza del 95 % el multiplicador t es $t_{12,3,0,025} = 2,173$. Sin utilizar un computador, podríamos redondear los grados de libertad a 12, en cuyo caso el multiplicador t es 2,179. Este cambio de 12,3 a 12 grados de libertad tiene poco efecto en el resultado final. La fórmula del intervalo de confianza da

$$(75,7 - 63,4) \pm (2,173)(4,14)$$

o

$$12,3 \pm 9,0$$

y el intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ es

$$(3,3, 21,3)$$

Tabla 6.7.2 Peso del tórax (mg)		
	Machos	Hembras
	67	73
	73	54
	85	61
	84	63
	78	66
	63	57
	80	75
		58
n	7	8
$\bar{y}$	75,7	63,4
s	8,4	7,5
ET	3,2	2,7

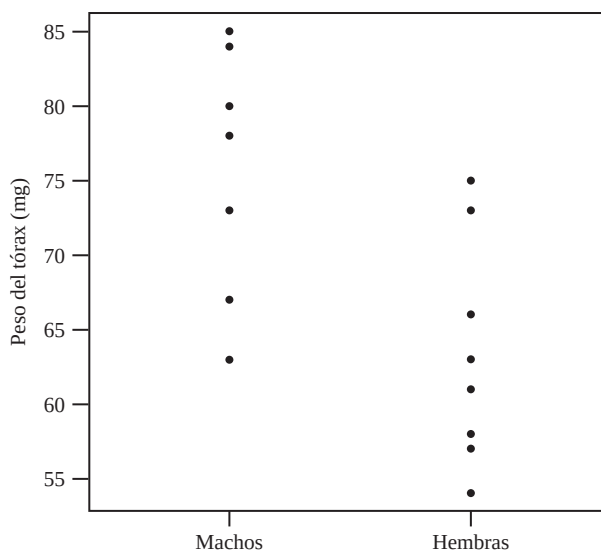


Figura 6.7.2 Presentación en paralelo de diagramas de puntos de los pesos del tórax

De acuerdo con el intervalo de confianza, podemos tener una confianza del 95 % de que la media poblacional del peso del tórax de mariposas Monarch macho (μ_1) sea mayor que el de las hembras (μ_2) en una cantidad que podría ser tan pequeña como 3,3 mg o tan grande como 21,3 mg.

Asimismo, para un intervalo de confianza del 90 % el multiplicador t es $t_{12,3,0,05} = 1,779$. La fórmula del intervalo de confianza da

$$(75,7 - 63,4) \pm (1,779)(4,14)$$

o

$$12,3 \pm 7,4$$

y el intervalo de confianza del 90 % para $(\mu_1 - \mu_2)$ es

$$(4,9, 19,7)$$

De acuerdo con el intervalo de confianza, podemos tener una confianza del 90 % de que la media poblacional del peso del tórax de mariposas Monarch macho (μ_1) sea mayor que el de las hembras (μ_2) en una cantidad que podría ser tan pequeña como 4,9 mg o tan grande como 19,7 mg. ■

Condiciones de validez En la Sección 6.5 planteamos las condiciones que hacían válido un intervalo de confianza para la media: se requiere que los datos se puedan ver como (1) una muestra aleatoria de (2) una población normal. Asimismo, cuando se comparan dos medias, se requieren dos muestras aleatorias independientes de poblaciones normales. Si los tamaños de la muestra son grandes, entonces la condición de normalidad no es crucial (debido al Teorema Central del Límite).

Ejercicios 6.7.1-6.7.14

6.7.1 En la Tabla 6.6.3, se presentaron datos de un experimento que comparaba dos tipos de cirugía. La puntuación de dolor media de los 49 niños que habían recibido tonsilectomía convencional era de 4,3, con una DT de 2,8. Para los 52 niños del grupo de coblación, la media era de 1,9 con una DT de 1,8. Utilice estos datos para construir un intervalo de confianza del 95 % para la diferencia de las medias poblacionales de las puntuaciones de dolor. [Nota: la fórmula (6.7.1) da un valor de 81,1 grados de libertad para estos datos].

6.7.2 El ácido ferúlico es un compuesto que puede tener un papel en la resistencia del maíz a las enfermedades. Un botánico midió la concentración de ácido ferúlico soluble en plantas de maíz criadas en semillero en oscuridad o en fotoperiodos de luz/oscuridad. Los resultados (nmol de ácido por g de tejido) fueron los que se muestran en la tabla⁴⁰.

	Oscuridad	Fotoperiodo
n	4	4
$\bar{y}$	92	115
s	13	13

- (a) Construya un intervalo de confianza del 95 % para la diferencia en concentración de ácido ferúlico bajo las dos condiciones de iluminación. (Suponga que las dos poblaciones de las que provienen los datos están distribuidas normalmente). [Nota: la fórmula (6.7.1) da un valor de 6 grados de libertad para estos datos].
- (b) Repita el apartado (a) para un nivel de confianza del 90 %.

6.7.3 (Continuación del 6.7.2) Utilizando su trabajo del Ejercicio 6.7.2(a) rellene el espacio en blanco: «tenemos una confianza del 95 % de que la diferencia de medias poblacionales sea al menos de _____ nmol/g».

6.7.4 Se realizó un estudio para determinar si el entrenamiento en relajación, ayudado con biorrealimentación y meditación, podría ayudar a reducir la presión sanguínea alta. Se asignaron sujetos aleatoriamente a un grupo de biorrealimentación y a un grupo de control. El grupo de biorrealimentación recibió entrenamiento durante ocho semanas. La tabla muestra la reducción en la presión sanguínea sistólica (mmHg) después de ocho semanas⁴¹. [Nota: la fórmula (6.7.1) da un valor de 190 grados de libertad para estos datos].

- (a) Construya un intervalo de confianza del 95 % para la diferencia en la respuesta media.
- (b) Interprete el intervalo de confianza del apartado (a) en el contexto de este estudio.

	Biorrealimentación	Control
n	99	93
$\bar{y}$	13,8	4,0
ET	1,34	1,30

6.7.5 Considere los datos del Ejercicio 6.7.4. Suponga que nos tememos que los datos de la presión sanguínea no provengan de distribuciones normales. ¿Significa esto que el

intervalo de confianza obtenido en el Ejercicio 6.7.4 no es válido? ¿Por qué o por qué no?

6.7.6 El tiempo de protrombina es una medida de la capacidad de coagulación de la sangre. Para 10 ratas tratadas con un antibiótico y 10 ratas de control, los tiempos de protrombina (en segundos) fueron los siguientes⁴²:

	Antibiótico	Control
n	10	10
$\bar{y}$	25	23
s	10	8

- Construya un intervalo de confianza del 90 % para la diferencia de las medias poblacionales. (Suponga que las dos poblaciones de las que provienen los datos están distribuidas normalmente). [Nota: la fórmula (6.7.1) da un valor de 17,2 grados de libertad para estos datos].
- ¿Por qué es importante suponer que las dos poblaciones están normalmente distribuidas en el apartado (a)?
- Interprete el intervalo de confianza del apartado (a) en el contexto de este estudio.

6.7.7 La tabla que acompaña a este ejercicio resume el consumo de sacarosa (mg en 30 minutos) de moscardones negros inyectados con Pargilina o con una solución salina (control)⁴³.

	Control	Pargilina
n	900	905
$\bar{y}$	14,9	46,5
s	5,4	11,7

- Construya un intervalo de confianza del 95 % para la diferencia en las medias poblacionales. [Nota: la fórmula (6.7.1) da un valor de 1.274 grados de libertad para estos datos].
- Repita el apartado (a) utilizando el nivel de confianza del 99 %.

6.7.8 En un estudio de campo sobre el comportamiento de apareamiento del grillo Mormon (*Anabrus simplex*), un biólogo advirtió que algunas hembras se apareaban con éxito mientras que otras eran rechazadas por los machos antes de completar el apareamiento. Surgió la cuestión de si algún aspecto del tamaño corporal podría tener un papel en el éxito del apareamiento. La tabla que acompaña a este ejercicio resume las medidas de la anchura de la cabeza (mm) en los dos grupos de hembras⁴⁴.

- Construya un intervalo de confianza del 95 % para la diferencia de las medias poblacionales. [Nota: la fórmula (6.7.1) da un valor de 35,7 grados de libertad para estos datos].
- Interprete el intervalo de confianza del apartado (a) en el contexto de este estudio.

- Utilizando su intervalo calculado en el apartado (a) para apoyar su respuesta, ¿existe evidencia fuerte de que la media poblacional de la altura de las cabezas sea mayor para las hembras con éxito que para las hembras sin éxito?

	Con éxito	Sin éxito
n	22	17
$\bar{y}$	8,498	8,440
s	0,283	0,262

6.7.9 En un experimento para evaluar el efecto de la dieta en la presión sanguínea, 154 adultos siguieron una dieta rica en frutas y vegetales. Un segundo grupo de 154 adultos siguió una dieta estándar. Se midieron al principio del estudio las presiones sanguíneas de los 308 sujetos. Ocho semanas más tarde, se midió de nuevo la presión sanguínea de los sujetos y se anotó para cada persona el cambio en dicha presión sanguínea. Los sujetos del grupo que seguía la dieta de frutas y vegetales tuvieron un descenso medio de la presión sanguínea sistólica de 2,8 mmHg más que los sujetos que seguían una dieta estándar. Un intervalo de confianza del 97,5 % de la diferencia entre las dos medias poblacionales es (0,9, 4,7)⁴⁵. Interprete el intervalo de confianza. Es decir, explique lo que significan los números del intervalo. (Véanse los Ejemplos 6.7.1 y 6.7.3).

6.7.10 Considere el experimento descrito en el Ejercicio 6.7.9. Para los mismos sujetos, el cambio en la presión sanguínea diastólica fue de 1,1 mmHg más grande, en promedio, para los sujetos que seguían la dieta de frutas y vegetales que para los sujetos que seguía una dieta estándar. Un intervalo de confianza del 97,5 % de la diferencia entre las dos medias poblacionales es (−0,3, 2,4). Interprete el intervalo de confianza. Es decir, explique lo que significan los números del intervalo. (Véanse los Ejemplos 6.7.1 y 6.7.3).

6.7.11 Unos investigadores estaban interesados en el efecto a corto plazo que la cafeína tenía en la frecuencia cardíaca. Reclutaron un grupo de voluntarios y midieron la frecuencia cardíaca en reposo de cada persona. Después hicieron que cada sujeto bebiera 6 onzas de café. 9 de los sujetos recibieron café que contenía cafeína y 11 café descafeinado. Después de 10 minutos se midió de nuevo la frecuencia cardíaca de cada persona. Los datos de la tabla muestran el cambio en la frecuencia cardíaca; un número positivo significa que la frecuencia cardíaca aumentó y un número negativo significa que la frecuencia cardíaca disminuyó⁴⁶.

	Cafeína	Descafeinado
	28	26
	11	1
	−3	0
	14	−4

(Continúa)

	Cafeína	Descafeinado
	-2	-4
	-4	14
	18	16
	2	8
	2	0
		18
		-10
n	9	11
$\bar{y}$	7,3	5,9
s	11,1	11,2
ET	3,7	3,4

- (a) Utilice estos datos para construir un intervalo de confianza del 90 % para la diferencia en el efecto medio que el café con cafeína tiene en la frecuencia cardíaca, en comparación con el café descafeinado. [Nota: la fórmula (6.7.1) da un valor de 17,3 grados de libertad para estos datos].
- (b) Utilizando el intervalo de confianza calculado en el apartado (a) para justificar su respuesta, ¿es razonable creer que la cafeína puede no afectar a la frecuencia cardíaca?
- (c) Utilizando el intervalo de confianza calculado en el apartado (a) para justificar su respuesta, ¿es razonable creer que la cafeína puede afectar a la frecuencia cardíaca? ¿Si es así, en cuánto?
- (d) ¿Son sus respuestas en (b) y (c) contradictorias? Explique su respuesta.

6.7.12 Considere los datos del Ejercicio 6.7.11. Dado que hay solo un pequeño número de observaciones en cada grupo, el intervalo de confianza calculado en dicho ejercicio solo es válido si las poblaciones subyacentes están normalmente distribuidas. ¿Es la condición de normalidad razonable en este caso? Apoye su respuesta con las gráficas apropiadas.

6.7.13 Un investigador estudió el efecto de la luz verde, en comparación con la luz roja, en la velocidad de crecimiento de plantas de judía. La tabla que se presenta a continuación muestra datos de las alturas de plantas (en pulgadas) desde el suelo hasta la primera bifurcación del tallo, dos semanas después de la germinación⁴⁷. Utilice estos datos para construir un intervalo de confianza del 95 % para la diferencia en el efecto medio que la luz roja tiene sobre el crecimiento de

las plantas de judía, en comparación con la luz verde. [Nota: la Fórmula (6.7.1) da un valor de 38 grados de libertad para estos datos].

	Rojo	Verde
	8,4	8,6
	8,4	5,9
	10,0	4,6
	8,8	9,1
	7,1	9,8
	9,4	10,1
	8,8	6,0
	4,3	10,4
	9,0	10,8
	8,4	9,6
	7,1	10,5
	9,6	9,0
	9,3	8,6
	8,6	10,5
	6,1	9,9
	8,4	11,1
	10,4	5,5
		8,2
		8,3
		10,0
		8,7
		9,8
		9,5
		11,0
		8,0
n	17	25
$\bar{y}$	8,36	8,94
s	1,50	1,78
ET	0,36	0,36

6.7.14 La distribución de los datos del Ejercicio 6.7.13 está algo sesgada, particularmente en el grupo rojo. ¿Significa esto que el intervalo de confianza calculado en el Ejercicio 6.7.13 no es válido? ¿Por qué o por qué no?

6.8 Perspectiva y resumen

En esta sección situaremos en perspectiva al Capítulo 6 relacionándolo con otros capítulos y también con otros métodos para analizar una única muestra de datos. Presentamos también un resumen condensado de los métodos presentados en este capítulo.

Distribuciones muestrales y análisis de datos

La teoría de la distribución muestral de $\bar{Y}$ (Sección 5.3) parecía requerir el conocimiento de valores (μ y σ) que en la práctica son desconocidos. En el Capítulo 6, sin embargo, hemos visto cómo realizar una inferencia sobre μ y sobre $(\mu_1 - \mu_2)$, incluyendo una evaluación de la precisión de dicha inferencia, utilizando solo la información proporcionada por la muestra. Por tanto, la teoría de las distribuciones muestrales nos ha llevado a un método práctico de análisis de datos.

En capítulos posteriores estudiaremos métodos más complejos de análisis de datos. Cada método se obtiene a partir de una distribución muestral apropiada. En la mayoría de los casos, sin embargo, no estudiaremos en detalle la distribución muestral.

Selección del nivel de confianza

Al ilustrar los métodos del intervalo de confianza, a menudo hemos escogido un nivel de confianza igual al 95 %. Sin embargo, debe recordarse que el nivel de confianza es arbitrario. Es verdad que en la práctica el nivel de confianza del 95 % es el más ampliamente utilizado. Sin embargo, no hay nada erróneo en utilizar un intervalo de confianza del 80 %, por ejemplo.

Características de otras medidas

Este capítulo ha presentado principalmente la estimación de una media poblacional, μ , y de una diferencia de medias poblacionales, $(\mu_1 - \mu_2)$. En algunas situaciones, puede desearse estimar otros parámetros de una población como la proporción poblacional (que consideraremos en el Capítulo 9). Los métodos de este capítulo se pueden extender a situaciones incluso más complejas. Por ejemplo, al evaluar una técnica de medida, el interés puede centrarse en la repetitibilidad de la técnica, indicado por la desviación típica de determinaciones repetidas. Como otro ejemplo, al definir los límites saludables, un investigador médico podría desear estimar el percentil 95 de los niveles de colesterol en suero en una cierta población. Lo mismo que la precisión de la media se puede indicar mediante un error típico o un intervalo de confianza, las técnicas estadísticas están también disponibles para especificar la precisión de los parámetros de la estimación tales como la desviación típica poblacional o el percentil 95.

Resumen de métodos de estimación

Para tener una referencia conveniente, resumimos en el siguiente cuadro los métodos de cálculo de intervalos de confianza presentados en este capítulo.

Error típico de la media

$$ET_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

Intervalo de confianza para μ

$$\text{Intervalo de confianza del 95 \% : } \bar{y} \pm t_{0,025} ET_{\bar{Y}}$$

Valor crítico $t_{0,025}$ obtenido de la distribución t de Student con $gl = n - 1$.

Los intervalos con otros niveles de confianza (como el 90 %, 99 %, etc.) se construyen de forma análoga (utilizando $t_{0,05}$, $t_{0,005}$, etc.).

La fórmula del intervalo de confianza es válida si (1) los datos se pueden considerar como una muestra aleatoria de una población grande, (2) las observaciones son independientes y (3) la población es normal. Si n es grande, entonces la condición (3) es menos importante.

Error típico de $(\bar{y}_1 - \bar{y}_2)$

$$ET_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{ET_1^2 + ET_2^2}$$

Intervalo de confianza para $(\mu_1 - \mu_2)$

Intervalo de confianza del 95 %:

$$(\bar{y}_1 - \bar{y}_2) \pm t_{0,025} ET_{(\bar{y}_1 - \bar{y}_2)}$$

Valor crítico $t_{0,025}$ obtenido de la distribución t de Student con

$$gl = \frac{(ET_1^2 + ET_2^2)^2}{ET_1^4/(n_1 - 1) + ET_2^4/(n_2 - 1)}$$

siendo $ET_1 = s_1/\sqrt{n_1}$ y $ET_2 = s_2/\sqrt{n_2}$.

Los intervalos con otros niveles de confianza (como el 90 %, 99 %, etc.) se construyen de forma análoga (utilizando $t_{0,05}$, $t_{0,005}$, etc.).

La fórmula del intervalo de confianza es válida si (1) los datos se pueden considerar como dos muestras aleatorias elegidas independientemente, (2) las observaciones son independientes dentro de cada muestra y (3) las dos poblaciones tienen distribución normal. Si n es grande, entonces la condición (3) es menos importante.

Ejercicios suplementarios 6.S.1-6.S.20

6.S.1 Para estudiar la conversión de nitrito a nitrato en la sangre, unos investigadores inyectaron a cuatro conejos una solución de moléculas de nitrito marcadas radiactivamente. Diez minutos después de la inyección, midieron para cada conejo el porcentaje de nitrito que se había convertido en nitrato. Los resultados fueron los siguientes⁴⁸:

51,1 55,4 48,0 49,5

- Para estos datos, calcule la media, la desviación típica y el error típico de la media.
- Construya un intervalo de confianza del 95 % para la media poblacional del porcentaje.
- Sin realizar ningún cálculo, ¿sería un intervalo de confianza del 99 % más ancho, más estrecho o de la misma anchura que el intervalo de confianza obtenido en el apartado (b)? ¿Por qué?

6.S.2 El diámetro del tallo de una planta de trigo es una característica importante debido a su relación con la rotura del tallo, que interfiere en la recolección de la cosecha. Un agrónomo midió el diámetro del tallo de ocho plantas de trigo rojo suave del cultivo *Tetrastichon*. Todas las observaciones se hicieron después de tres semanas de la floración de la planta. Los diámetros del tallo (mm) fueron los siguientes⁴⁹:

2,3 2,6 2,4 2,2 2,3 2,5 1,9 2,0

La media de estos datos es 2,275 y la desviación típica 0,238.

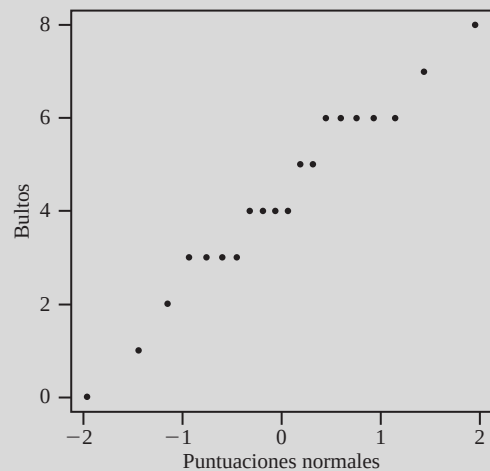
- Calcule el error típico de la media.
- Construya un intervalo de confianza del 95 % para la media poblacional.
- Defina con palabras la media poblacional que estimó en el apartado (b). (Véase el Ejemplo 6.1.1).

6.S.3 En relación con el Ejercicio 6.S.2.

- ¿Qué condiciones son necesarias para que el intervalo de confianza sea válido?
- ¿Se cumplen esas condiciones? ¿Cómo lo sabe?
- ¿Cuál de estas condiciones es la más importante?

6.S.4 En referencia al Ejercicio 6.S.2, suponga que los datos de las ocho plantas se consideran un estudio piloto, y que el agrónomo quiere ahora diseñar un nuevo estudio en el que desea que el error típico de la medida sea solo de 0,03 mm. ¿Cuántas plantas deberían medirse en el nuevo estudio?

6.S.5 Una muestra de 20 larvas de la mosca de la fruta (*Drosophila melanogaster*) se incubó a 37 °C durante 30 minutos. Existe la teoría de que esa exposición al calor causa que los cromosomas politénicos localizados en las glándulas salivales de la mosca se desarrollen, creando bultos en el brazo del cromosoma que son visibles al microscopio. La siguiente gráfica de probabilidad normal apoya el uso de una curva normal para modelar la distribución de los bultos⁵⁰.



El número medio de bultos en las 20 observaciones fue de 4,30, con una desviación típica de 2,03.

- (a) Construya un intervalo de confianza del 95 % para μ .
- (b) En el contexto de este problema, describa lo que representa μ . Es decir, ¿de qué cantidad es el intervalo de confianza del apartado (a)?
- (c) La gráfica de probabilidad normal muestra que los puntos se alinean en bandas horizontales. ¿Es esta clase de comportamiento sorprendente en este tipo de datos? Explique su respuesta.

6.S.6 En un periodo de aproximadamente nueve meses, 1.353 mujeres informaron sobre la duración de sus ciclos menstruales. Para el primer ciclo de cada mujer, el tiempo medio del ciclo fue de 28,86 días, y la desviación típica de los 1.353 tiempos fue de 4,24 días⁵¹.

- (a) Construya un intervalo de confianza del 99 % para la media poblacional del tiempo de ciclo.
- (b) Como los ritmos ambientales pueden influir en los ritmos biológicos, se podría hacer la hipótesis de que la media poblacional del tiempo de ciclo menstrual es de 29,5 días, la longitud del mes lunar. ¿Es el intervalo de confianza del apartado (a) consistente con esta hipótesis?

6.S.7 En referencia a los datos sobre el ciclo menstrual del Ejercicio 6.S.6.

- (a) Durante todo el periodo de tiempo del estudio, las mujeres informaron que un total de 12.247 ciclos. Cuando se incluyen todos estos ciclos, el tiempo medio del ciclo es de 28,22 días. Explique por qué se podría esperar que esta media fuera más pequeña que el valor de 28,86 dado en el Ejercicio 6.S.6. (*Sugerencia*: si cada mujer informara un tiempo fijo de ciclo, ¿qué mujeres contribuirían con más ciclos al total de las 12.247 observaciones?).
- (b) En lugar de utilizar solo el primer ciclo como en el Ejercicio 6.S.6, se podrían utilizar los primeros cuatro ciclos de cada mujer, obteniéndose así $1.353 \times 4 = 5.412$ observaciones. Se podría entonces calcular la media y la desviación típica de las 5.412 observaciones y dividir la DT por $\sqrt{5412}$ para obtener el ET. Esto produciría un valor del ET mucho menor que el obtenido en el Ejercicio 6.S.6. ¿Por qué este planteamiento no es válido?

6.S.8 Para los pesos a nacer de los 28 corderos del Ejemplo 6.2.2, la media es 5,1679 kilogramos, la DT es 0,6544 kg y el ET es 0,1237 kg.

- (a) Construya un intervalo de confianza del 95 % para la media poblacional.
- (b) Construya un intervalo de confianza del 99 % para la media poblacional.
- (c) Interprete el intervalo de confianza obtenido en el apartado (a). Es decir, explique lo que significan los números del intervalo. (*Sugerencia*: véanse los Ejemplos 6.3.4 y 6.3.5).
- (d) A menudo, los investigadores resumen sus datos en informes y artículos escribiendo $\bar{y} \pm DT(5,17 \pm 0,65)$ o $\bar{y} \pm ET(5,17 \pm 0,12)$. Si la investigadora de este estudio planea comparar el peso medio al nacer de estos corde-

ros Rambouillet con los de otra raza, Booroolas, ¿qué estilo de presentación debería utilizar?

6.S.9 En referencia al Ejercicio 6.S.8.

- (a) ¿Qué condiciones se requieren para la validez de los intervalos de confianza?
- (b) ¿Cuál de las condiciones del apartado (a) se puede comprobar (aproximadamente) a partir del programa de la Figura 6.2.1?
- (c) Los nacimientos de gemelos se excluyeron de los datos de peso al nacer de los corderos. Si se hubieran incluido los nacimientos de gemelos, ¿sería válido el intervalo de confianza? ¿Por qué o por qué no?

6.S.10 Unos investigadores midieron el número de especies de árbol en 69 parcelas vegetales en el Lama Forest de Benin, oeste de África⁵². El número de especies varió desde un mínimo de 1 hasta un máximo de 12. La media muestral fue de 6,8 y la DT muestral fue de 2,4, lo que produce un intervalo de confianza del 95 % de (6,2, 7,4). Sin embargo, el número de especies de árboles en una gráfica toma solo valores enteros. ¿Significa esto que intervalo de confianza debería ser (7, 7)? ¿O significa que deberíamos redondear los extremos del intervalo de confianza y dejarlo como (6, 7)? ¿O debería el intervalo de confianza ser realmente (6,2, 7,4)? Explique su respuesta.

6.S.11 Como parte de un estudio de la variación natural en la química de la sangre, se midieron las concentraciones de potasio en suero en 84 mujeres con buena salud. La concentración media fue de 4,36 mEq/l y la desviación típica fue de 0,42 mEq/l. La tabla siguiente presenta una distribución de frecuencias de los datos⁵³.

Potasio en suero (mEq/l)	Número de mujeres
[3,1, 3,4)	1
[3,4, 3,7)	2
[3,7, 4,0)	7
[4,0, 4,3)	22
[4,3, 4,6)	28
[4,6, 4,9)	16
[4,9 5,2)	4
[5,2, 5,5)	3
[5,5, 5,8)	1
Total	84

- (a) Calcule el error típico de la media.
- (b) Construya un histograma de los datos e indique sobre dicho histograma los intervalos $\bar{y} \pm DT$ e $\bar{y} \pm ET$. (Véase la Figura 6.2.1).
- (c) Construya un intervalo de confianza del 95 % para la media poblacional.

- (d) Interprete el intervalo de confianza obtenido en el apartado (c). Es decir, explique lo que significan los números del intervalo. (*Sugerencia:* véanse los Ejemplos 6.3.4 y 6.3.5).

6.S.12 En referencia al Ejercicio 6.S.11, los médicos, en sus diagnósticos, utilizan a menudo el término «límites de referencia» para juzgar los valores químicos de la sangre. Son límites entre los que se espera encontrar al 95 % de la gente con buena salud. ¿Sería un intervalo de confianza del 95 % para la media una elección razonable de «límites de referencia» para el potasio en suero en mujeres? ¿Por qué o por qué no?

6.S.13 En referencia al Ejercicio 6.S.11, suponga que se realiza un estudio similar al año siguiente, para incluir las medidas de potasio en suero de 200 mujeres con buena salud. Basándose en los datos del Ejercicio 6.S.11, indique lo que podría predecir sobre el valor de:

- (a) La DT de las nuevas medidas.
(b) El ET de las nuevas medidas.

6.S.14 Un agrónomo seleccionó aleatoriamente seis plantas de trigo de una parcela y después, para cada planta, seleccionó 12 semillas de la parte principal de la espiga. Pesándolas, secándolas y volviéndolas a pesar, se terminó el porcentaje de humedad en cada grupo de semillas. Los resultados fueron los siguientes⁵⁴:

62,7 63,6 60,9 63,0 62,7 63,7

- (a) Calcule la media, la desviación típica y el error típico de la media.
(b) Construya un intervalo de confianza del 90 % para la media poblacional.

6.S.15 Como parte del Estudio Nacional sobre Salud y Nutrición, se comprobaron los niveles de hemoglobina de una muestra de 1.139 hombres de 70 o más años de edad⁵⁵. La media muestral fue de 145,3 g/l y la desviación típica fue de 12,87 g/l.

- (a) Utilice estos datos para construir un intervalo de confianza de 95 % para μ .
(b) ¿Proporciona el intervalo de confianza del apartado (a) límites en los que podemos esperar que estén el 95 % de los datos de la muestra? ¿Por qué o por qué no?
(c) ¿Proporciona el intervalo de confianza del apartado (a) límites en los que podemos esperar que estén el 95 % de los datos de la población? ¿Por qué o por qué no?

6.S.16 Los datos siguientes corresponden a los valores durante 16 semanas de coliformes fecales (MPN/100 ml) en Dairy Creek en San Luis Obispo County, California⁵⁶.

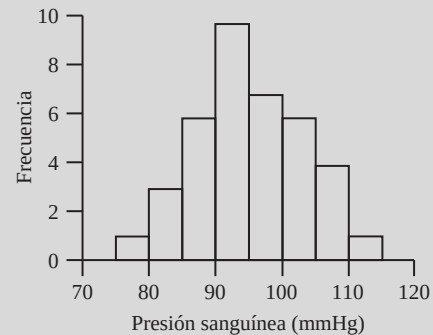
203 215 240 236 217 296 301 190
197 203 210 215 270 290 310 287

- (a) Valores por encima de 225 MPN/100 ml se consideran inseguros. ¿Qué tipo de intervalo unilateral (límite superior o límite inferior) sería apropiado para evaluar la seguridad de este arroyo? Explique su razonamiento.

- (b) Utilizando una confianza de 95 %, construya el intervalo elegido en el apartado (a).

- (c) Basándose en su intervalo del apartado (b), ¿qué conclusiones puede extraer con respecto a la seguridad del agua?

6.S.17 Se midió la presión sanguínea (promedio de las medidas sistólica y diastólica) de 38 personas⁵⁷. El promedio fue de 94,5 (mmHg). Se muestra a continuación un histograma de los datos.



¿Cuál de los siguientes es aproximadamente un intervalo de confianza de 95 % para la media poblacional de la presión sanguínea? Explique su respuesta.

- (i) $94,5 \pm 16$ (ii) $94,5 \pm 8$
(iii) $94,5 \pm 2,6$ (iv) $94,5 \pm 1,3$

6.S.18 Suponga que desea estimar la presión sanguínea media de estudiantes de su universidad con una precisión de 2 mmHg, con una confianza del 95 %.

- (a) Utilizando los datos que se muestran en el Ejercicio 6.S.17 como datos piloto de su estudio, determine el tamaño de la muestra (aproximado) necesario para conseguir su objetivo. (*Sugerencia:* necesitará utilizar la gráfica para hacer algunas estimaciones visuales).
(b) Suponga que su universidad es una pequeña escuela privada que solo tiene 500 estudiantes. ¿Sería válido el intervalo basado en su tamaño de la muestra? Explique su respuesta. ¿Piensa que sería demasiado ancho o demasiado estrecho?

6.S.19 Se sabe que el consumo de alcohol durante el embarazo puede dañar al feto. Para estudiar este fenómeno, 10 ratones hembra embarazadas recibieron una baja dosis de alcohol. Cuando cada una de ellas dio a luz, se midió el peso al nacer de cada cría. Suponga que los ratones hembra dieron a luz un total de 85 crías, de manera que el experimentador tenía 85 observaciones de Y = peso al nacer. Para calcular el error típico de la media de estas 85 observaciones, el investigador podría calcular la desviación típica de dichas 85 crías y dividir por $\sqrt{85}$. ¿Sobre qué base se podría plantear una objeción a este método de calcular el ET?

6.S.20 ¿Es precisa la información nutricional de los alimentos producidos comercialmente? En un estudio, unos investigadores muestrearon 13 paquetes de pollo congelado bajo en calorías que informaban de un contenido de 252 calorías por paquete. El valor medio de las calorías de los paquetes muestreados fue de 306 con una desviación típica muestral de 51 calorías⁵⁸.

(a) Calcule un intervalo de confianza del 95 % para la media poblacional del contenido en calorías de los congelados de pollo.

(b) Basándose en el intervalo calculado en el apartado (a), ¿qué piensa sobre el contenido en calorías indicado en cada paquete?

(c) Los fabricantes pueden ser sancionados si proporcionan *menos* alimento que el que anuncia. ¿Cómo se relaciona este hecho con sus resultados en los apartados (a) y (b)?

Notas

1. Datos proporcionados por Dennis Frey, California Polytechnic State University, San Luis Obispo.

2. Newman, S., Everson, D. O., Gunsett, F. C. y Christian, R. E. (1984). Analysis of two-and three-way crosses among Ramhouillet, Targhee, Columbia, and Suffolk sheep for three preweaning traits. Manuscrito no publicado. Datos iniciales cortesía de S. Newman.

3. Adaptado de los artículos siguientes. Potkin, S. G., Cannon, H. E., Murphy, D. L. y Wyatt, R. J. (1978). Are paranoid schizophrenics biologically different from other schizophrenics? *New England Journal of Medicine* **298**, 61-66. Murphy, D. L., Wright, C., Buchsbaum, M., Nichols, A., Costa, J. L. y Wyatt, R. J. (1976). Platelet and plasma amine oxidase activity in 680 normals: Sex and age differences and stability over time. *Biochemical Medicine* **16**, 254-265. Los datos representados son ficticios pero realistas, reconstruidos a partir de los histogramas e información resumen dados por Potkin et al. y Murphy et al.

4. Basado en los datos expuestos en Rea, T. M., Nash, J. F., Zabik, J. E., Born, G. S. y Kessler, W. V. (1984). Effects of toluene inhalation on brain biogenic amines in the rat. *Toxicology* **31**, 143-150.

5. Basado en un experimento de M. Morales.

6. Adaptado de Cherney, J. H., Volenec, J. J. y Nyquist, W. E. (1985). Sequential fiber analysis of forage as influenced by sample weight. *Crop Science* **25**, n.º 6 (nov./dic. 1985), 1113-1115 (Tabla 1). Con permiso de la Crop Science Society of America, Inc. Datos iniciales cortesía de W. E. Nyquist.

7. Dice, L. R. (1932). Variation in the geographic race of the deermouse, *Peromyscus maniculatus bairdii*. *Occasional Papers of the Museum of Zoology, University of Michigan*, n.º 239. Datos reproducidos en Simpson, G. G., Roe, A. y Lewontin, R. C. (1960). *Quantitative Zoology*. New York: Harcourt, Brace, p. 79.

8. Bodor, N. y Simpkins, J. W. (1983). Redox delivery system for brain-specific, sustained release of dopamine. *Science* **221**, 65-67.

9. Student (W. S. Gosset) (1908). The probable error of a mean. *Biometrika* **6**, 1-25.

10. The Writing Group for the PEPI Trial (1996). Effects of hormone therapy on bone mineral density. *Journal of the American Medical Association* **276**, 1389-1396. En este estudio se comparó el cambio en la densidad mineral del hueso durante 36 meses para cuatro medicaciones y un placebo. (La densidad mineral del hueso de la cadera se midió al comienzo del experimento y 36 meses después). Solo los datos de aquellas mujeres que siguieron el protocolo experimental se usaron en el ejemplo. Las desviaciones típicas se calculan basándose en los errores típicos indicados en el artículo.

11. Datos recogidos por Denise D'Abundo, Oberlin College, abril 1991.

12. Bockman, D. E. y Kirby, M. L. (1984). Dependence of thymus development on derivatives of the neural crest. *Science* **223**, 498-500. Copyright 1984 de la AAAS.

13. Brown, S. A., Riviere, J. E., Coppoc, G. L., Hinsman, E. J., Carlton, W. W. y Steckel, R. R. (1985). Single intravenous and multiple intramuscular dose pharmacokinetics and tissue residue profile of gentamicin in sheep. *American Journal of Veterinary Research* **46**, 69-74. Datos iniciales cortesía de S. A. Brown y G. L. Coppoc.

14. Lobstein, D. D. (1983). A multivariate study of exercise training effects on beta-endorphin and emotionality in psychologically normal, medically healthy men. Ph. D. thesis, Purdue University. Datos iniciales cortesía del autor.

15. Nicholson, R. L. y Moraes, W. B. C. (1980). Survival of *Colletotrichum graminicola*: Importance of the spore matrix. *Phytopathology* **70**, 255-261.

16. Adaptado de Morris, J. G., Gripe, W. S., Chapman, H. L., Jr., Walker, D. F., Armstrong, J. B., Alexander, J. D., Jr., Miranda, R., Sanchez, A., Jr., Sanchez, B., Blair-West, J. R. y Denton, D. A. (1984). Selenium deficiency in cattle associated with Heinz bodies and anemia. *Science* **223**, 491-492. Copyright 1984 de la AAAS.

17. Shaffer, P. L. y Rock, G. C. (1983). Tufted apple budmoth (*Lepidoptera: Tortricidae*): Effects of constant daylengths and temperatures on larval growth rate and determination of larval-pupal ecdysis. *Environmental Entomology* **12**, 76-80.
18. Bishop, N. J., Morley, R., Day, J. P. y Lucas, A. L. (1997). Aluminum neurotoxicity in preterm infants receiving intravenous-feeding solutions. *New England Journal of Medicine* **336**, 1557-1561.
19. Kaufman, J. S., Reda, D. J., Fye, C. L., Goldfarb, D. S., Henderson, W. G., Kleinman, J. G. y Vaamonde, C. A. (1998). Subcutaneous compared with intravenous epoetin in patients receiving hemodialysis. *New England Journal of Medicine* **339**, 578-583.
20. Basado en datos proporcionados por C. H. Noller.
21. Esta es aproximadamente la DT de la población norteamericana de hombres de mediana edad. Véase Moore, F. E. y Gordon, T. (1973). Serum cholesterol levels in adults, United States 1960-62. *U.S. National Center for Health Statistics, Vital and Health Statistics*, Serie 11, n.º 22. Washington, D.C.: U.S. Department of Health, Education and Welfare.
22. Pappas, T. y Mitchell, C. A. (1984). Effects of seismic stress on the vegetative growth of *Glycine max* (L.) Merr. cv. Wells II. *Plant, Cell and Environment* **8**, 143-148.
23. Noll, S. L., Waibel, P. E., Cook, R. D. y Witmer, J. A. (1984). Biopotency of methionine sources for young turkeys. *Poultry Science* **63**, 2458-2470.
24. Schaeffer, J., Andrysiak, T. y Ungerleider, J. T. (1981). Cognition and long-term use of ganja (cannabis). *Science* **213**, 465-466.
25. Desai, R. (1982). Anatomical study of the canine male and female pelvic diaphragm and the effect of testosterone on the status of the levator ani of male dogs. *Journal of the American Animal Hospital Association* **18**, 195-202.
26. Nicholson, R. L. y Moraes, W. B. C. (1980). Survival of *Colletotrichum graminicola*: Importance of the spore matrix. *Phytopathology* **70**, 255-261. Datos iniciales cortesía de R. L. Nicholson.
27. Las probabilidades de la Tabla 6.5.2 se estimaron mediante simulación en computador realizada por M. Samuels y R. P. Becker. El error típico de cada estimación de probabilidad es menor que 0,0015. Las fuentes de las distribuciones de los padres se dan en las Notas 7 y 8 del Capítulo 5.
28. Burnett, A. y Haywood, A. (1997). A statistical analysis of differences in sediment yield over time on the West Branch of the Black River. Manuscrito no publicado, Oberlin College.
29. Hessell, E. A., Johnson, D. D., Ivey, T. D. y Miller, D. W. (1980). Membrane vs bubble oxygenator for cardiac operations. *Journal of Thoracic and Cardiovascular Surgery* **80**, 111-122.
30. Peters, H. G. y Bademan, H. (1963). The form and growth of stellate cells in the cortex of the guinea-pig. *Journal of Anatomy* (London) **97**, 111-117.
31. Kaneto, A., Kosaka, K. y Nakao, K. (1967). Effects of stimulation of the vagus nerve on insulin secretion. *Endocrinology* **80**, 530-536. Copyright © 1967 de la Endocrine Society.
32. Simmons, F. J. (1943). Occurrence of superparasitism in *Nemeritis canescens*. *Revue Canadienne de Biologie* **2**, 15-40. Datos reproducidos en Williams, C. B. (1964). *Patterns in the Balance of Nature*. London: Academic Press, p. 223.
33. Estos datos son diámetros a la altura del pecho de árboles sicomoros americanos en la pradera del Vermilion River. Datos recogidos por Emily Norland, Oberlin College, marzo de 1995.
34. Adaptado de Sanders, K. (2004). A quantitative, vegetative, and reproductive comparison of *Centromadia parryi* ssp. *congdonii* in two locations. Master's Thesis, California Polytechnic State University.
35. Hunter, A. y Terasaki, T. (1993). Statistical analysis comparing vital capacities of brass majors in the Conservatory and a normal population. Manuscrito no publicado, Oberlin College. Todos los sujetos eran hombres, entre 18 y 21 años de edad y con alturas entre 175 y 183 cm. Como la capacidad vital está relacionada con la altura, los datos iniciales se ajustaron ligeramente, utilizando regresión lineal, para controlar el efecto de la altura.
36. Chang, K. (2005). Randomized controlled trial of Coblation versus electrocautery tonsillectomy. *American Academy of Otolaryngology-Head and Neck Surgery* **132**, 273-280.
37. Knight, S. L. y Mitchell, C. A. (1983). Enhancement of lettuce yield by manipulation of light and nitrogen nutrition. *Journal of the American Society for Horticultural Science* **108**, 750-754. Datos iniciales cortesía de los autores. (Los tamaños muestrales reales eran iguales. Algunas observaciones del ejercicio se han omitido).
38. O'Marra, S. (1996). Antibacterial soaps: Myth or reality. Manuscrito no publicado, Oberlin College. El objetivo principal de este estudio era evaluar la efectividad de los jabones antibacterianos. Una solución de jabón antibacteriano mataba a los *E. coli*, a diferencia del jabón no antibacteriano y del control. La solución de jabón era una solución 1:4 de jabón en agua.
39. Ahern, T. (1998). Statistical analysis of EIN plants treated with ancyimidol and H₂O. Manuscrito no publicado, Oberlin College. En este experimento se utilizó la cepa mutante EIN (e-longated in-ternode) de *Brassica*. Los datos

presentados aquí son un subconjunto seleccionado aleatoriamente del conjunto completo de datos.

40. Hagerman, A. E. y Nicholson, R. L. (1982). High-performance liquid chromatographic determination of hydroxycinnamic acids in the maize mesocotyl. *Journal of Agricultural and Food Chemistry* **30**, 1098-1102. Reimpreso con permiso. Copyright 1982 American Chemical Society.
41. Patel, C., Marmot, M. M. y Terry, D. J. (1981). Controlled trial of biofeedback-aided behavioral methods in reducing mild hypertension. *British Medical Journal* **282**, 2005-2008.
42. Lipsky, J. J., Lewis, J. C. y Novick, W. J., Jr. (1984). Production of hypoprothrombinemia by Moxalactam and 1-methyl-5-thiotetrazole in rats. *Antimicrobial Agents and Chemotherapy* **25**, 380-381.
43. Long, T. F. y Murdock, L. L. (1983). Stimulation of blowfly feeding behavior by octopaminergic drugs. *Proceedings of the National Academy of Sciences* **80**, 4159-4163. Datos iniciales cortesía de los autores y de L. C. Sudlow.
44. Gwynne, D. T. (1981). Sexual difference theory: Mormon crickets show role reversal in mate choice. *Science* **213**, 779-780. Copyright 1981 de la AAAS. Datos proporcionados por cortesía del autor.
45. Appel, L. J., *et al.* (1997). A clinical trial of the effects of dietary patterns on blood pressure. *New England Journal of Medicine* **336**, 1117-1124.
46. Crawford, F. y Piper, Y. (1999). How does caffeine influence heart rate? Manuscrito no publicado, Oberlin College. Había 10 sujetos en el grupo de la cafeína, pero se detectó un *outlier* en los datos.
47. Gent, A. (1999). Datos no publicados recogidos en el Oberlin College. Los colores de la luz se crearon utilizando geles: trozos finos de plástico coloreado utilizados en las luces de teatro.
48. Parks, N. J., Krohn, K. A., Mathis, C. A., Chasko, J. H., Geiger, K. R., Gregor, M. E. y Peek, N. F. (1981). Nitrogen-13-labelled nitrite and nitrate: Distribution and metabolism after intratracheal administration. *Science* **212**, 58-61. Copyright 1981 de la AAAS. Datos iniciales cortesía de N. J. Parks.
49. Krick, J. A. (1982). Effects of seeding rate on culm diameter and the inheritance of culm diameter in soft red winter wheat (*Triticum aestivum* L. em Thell). Master's thesis, Department of Agronomy, Purdue University. Datos iniciales cortesía de J. A. Krick y H. W. Ohm. Cada diámetro es la media de medidas tomadas en seis localizaciones prescritas del tallo.
50. Datos recogidos por Deborah Ignatoff, Oberlin College, primavera de 1997.
51. Bailey, J. y Marshall, J. (1970). The relationship of the post-ovulatory phase of the menstrual cycle to total cycle length. *Journal of Biosocial Science* **2**, 123-132.
52. Nansen, C., Tchabi, A. y Meikle, W. G. (2001). Successional sequence of forest types in a disturbed dry forest reserve in southern Benin, West Africa. *Journal of Tropical Ecology* **17**, 525-539.
53. Datos no publicados cortesía de W. F. Jacobson.
54. Dale, E. M. y Housley, T. L. (1986). Sucrose synthase activity in developing wheat endosperms differing in maximum weight. *Plant Physiology* **82**, 7-10. Datos iniciales cortesía de los autores.
55. Véase la Nota 23 del Capítulo 3.
56. Adaptado de los datos cortesía de la Morro Bay National Estuary Foundation, 2009.
57. Gráfico creado de datos incluidos en Erne, P., Bolli, P., Buerigisser, E. y Buehler, F. R. (1984). Correlation of platelet calcium with blood pressure. *New England Journal of Medicine* **310**, 1084-1088. Reimpreso con permiso. Datos iniciales cortesía de F. R. Buehler.
58. Urban L. E., *et al.* (2010). The accuracy of stated energy contents of reduced-energy, commercially prepared foods. *J. Am. Diet Assoc.* **110**, 116-123.

COMPARACIÓN DE DOS MUESTRAS INDEPENDIENTES

Objetivos

En este capítulo continuaremos nuestro estudio de la comparación de dos muestras independientes presentando los contrastes de hipótesis. En particular:

- exploraremos cómo se puede utilizar la aleatorización para formar la base de una inferencia estadística;
- demostraremos cómo realizar un test t de dos muestras para comparar medias muestrales y explicaremos cómo se relaciona este texto con el intervalo de confianza de la diferencia de dos medias;
- presentaremos la interpretación del P valor;
- examinaremos más de cerca cómo la confusión y la relación engañosa puede limitar la utilidad de un estudio;
- compararemos la inferencia causal frente a asociativa y sus relaciones con los experimentos y estudios observacionales;
- presentaremos los conceptos de nivel de significación, tamaño del efecto, errores Tipo I y II, y potencia;
- distinguiremos entre tests direccionales y no direccionales y examinaremos cómo comparar los P valores de estos tests;
- consideraremos las condiciones bajo las que el uso de test t es válido;
- mostraremos cómo comparar distribuciones utilizando el test de Wilcoxon-Mann-Whitney.

7.1 Contraste de hipótesis: el test de aleatorización

Consideremos que se toma una muestra de una población y después se divide aleatoriamente dicha muestra en dos partes. Podríamos esperar que las dos partes de la muestra parecieran similares, pero no exactamente iguales. Supongamos ahora que tomamos muestras de dos poblaciones. Si las dos muestras parecen bastante similares entre sí, podríamos inferir que las dos poblaciones son idénticas. Si las muestras parecen bastante diferentes, podríamos inferir que las poblaciones son diferentes. La pregunta es: «¿Cómo deben ser de diferentes las dos muestras para inferir que las poblaciones que las generaron son realmente diferentes?».

Una forma de resolver esta cuestión es comparar las dos medias muestrales y ver cuánto se diferencian en comparación con la cantidad de diferencia que esperaríamos ver debida al azar*. El test de aleatorización nos proporciona una forma de medir la variabilidad en la diferencia de dos medias muestrales.

Ejemplo 7.1.1 Flexibilidad

Un investigador estudió la flexibilidad de siete mujeres, cuatro de las cuales estaban en una clase de aeróbic y las otras tres eran bailarinas. Una medida que tomó fue la «flexibilidad del tronco»: qué distancia podía cada una de las mujeres estirarse mientras estaba sentada en el suelo[†]. Las medidas (en centímetros) se muestran en la Tabla 7.1.1¹.

* Se podrían comparar las dos medianas muestrales en vez de las medias. Comparamos medias porque tenemos un proceso denominado test t , que se presentará en la sección siguiente y dicho test se basa en las medias.

[†]Estos datos son parte de un estudio mayor. Trabajamos solo con un subconjunto del estudio completo para simplificar.

Tabla 7.1.1	
Aerobic	Bailarinas
38	48
45	59
58	61
64	
media 51,25	56,00

¿Proporcionan estos datos evidencia de que la flexibilidad está asociada con ser bailarina?

Si ser bailarina no tiene efecto en la flexibilidad, entonces se podría argumentar que los siete puntos de datos del estudio provienen de una población como: algunas mujeres tienen mayor flexibilidad del tronco que otras, pero esto no tiene nada que ver con ser bailarina.

Otra forma de decir esto es:

Afirmación: Las siete medidas de flexibilidad del tronco provienen de una única población. Las etiquetas «aerobic» y «bailarinas» son arbitrarias y no tienen nada que ver con la flexibilidad medida como flexibilidad del tronco. ■

Si la afirmación indicada en el Ejemplo 7.1.1 es verdadera, entonces cualquier reordenación de las siete observaciones en dos grupos, con cuatro mujeres «aerobic» y tres mujeres «bailarinas» es igual de probable que cualquier otra reordenación. De hecho, podríamos imaginar que escribimos las siete observaciones en siete tarjetas, las barajamos, y después extraemos cuatro de ellas etiquetándolas del grupo «aerobic», y las otras tres como del grupo «bailarinas».

Ejemplo 7.1.2 Flexibilidad

Hay 35 posibles formas de dividir las medidas de flexibilidad del tronco de las siete observaciones en dos grupos de tamaños 4 y 3. La Tabla 7.1.2 muestra esas 35 posibilidades, junto con la diferencia en las medias muestrales de cada una. (Presentamos las medias con tres cifras decimales, ya que utilizaremos estos valores en cálculos futuros). Las dos muestras obtenidas en el estudio se indican en primer lugar, seguidas por las otras 34 formas en las que se podían haber obtenido las muestras.

La Figura 7.1.1 es una representación visual de estos 35 posibles valores. El resultado observado de $-4,75$, que está resaltado, no está muy lejos de la mitad de la distribución.

Supongamos que las etiquetas «aerobic» y «bailarinas» son, de hecho, arbitrarias y no tienen nada que ver con la flexibilidad del tronco. Entonces, cada uno de los 35 resultados que se presentan en la Tabla 7.1.2, y se muestran en la Figura 7.1.1, es igualmente probable. De las 35 diferencias, 20 son al menos tan grandes en módulo como el $-4,75$ obtenido en el estudio. Estas se muestran con tipo negrilla en la tabla y se rellenan con negro o gris en la figura. Por tanto, si la afirmación es cierta (que las etiquetas «aerobic» y «bailarinas» son arbitrarias), hay una probabilidad de $20/35$ de obtener una diferencia en las medias muestrales tan grande, en módulo, como la diferencia que ha sido observada.

La fracción $20/35$ es aproximadamente igual a $0,57$, que es más bien grande. Por tanto, los datos observados son consistentes con la afirmación de que las etiquetas «aerobic» y «bailarinas» son arbitrarias y no tienen nada que ver con la flexibilidad. Si la afirmación es cierta, esperaremos ver una diferencia en las medias muestrales de $4,75$ o más la mitad de las veces, únicamente por efecto del azar. Por tanto, estos datos proporcionan poca evidencia de que la flexibilidad esté asociada con el hecho de ser bailarina. ■

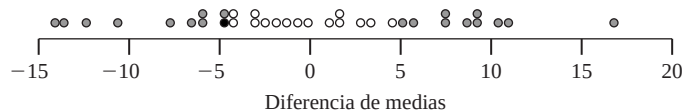


Figura 7.1.1 Distribución de los valores de «diferencia de medias», con el resultado observado de $-4,75$ coloreado en negro, y los valores con resultados observados tan extremos o más (en módulo) que $4,75$ coloreados en gris

Tabla 7.1.2				
Muestra 1 («aerobic»)	Muestra 2 («bailarinas»)	Media de la muestra 1	Media de la muestra 2	Diferencia de medias
38 45 58 64	48 59 61	51,25	56,00	– 4,75
38 45 58 48	64 59 61	47,25	61,33	– 14,08
38 45 58 59	64 48 61	50,00	57,67	– 7,67
38 45 58 61	64 48 59	50,50	57,00	– 6,50
38 45 64 48	58 59 61	48,75	59,33	– 10,58
38 45 64 59	58 48 61	51,50	55,67	– 4,17
38 45 64 61	58 48 59	52,00	55,00	– 3,00
38 45 48 59	58 64 61	47,50	61,00	– 13,50
38 45 48 61	58 64 59	48,00	60,33	– 12,33
38 45 59 61	58 64 48	50,75	56,67	– 5,92
38 58 64 48	45 59 61	52,00	55,00	– 3,00
38 58 64 59	45 48 61	54,75	51,33	3,42
38 58 64 61	45 48 59	55,25	50,67	4,58
38 58 48 59	45 64 61	50,75	56,67	– 5,92
38 58 48 61	45 64 59	51,25	56,00	– 4,75
38 58 59 61	45 64 48	54,00	52,33	1,67
38 64 48 59	45 58 61	52,25	54,67	– 2,42
38 64 48 61	45 58 59	52,75	54,00	– 1,25
38 64 59 61	45 58 48	55,50	50,33	5,17
38 48 59 61	45 58 64	51,50	55,67	– 4,17
45 58 64 48	38 59 61	53,75	52,67	1,08
45 58 64 59	38 48 61	56,50	49,00	7,50
45 58 64 61	38 48 59	57,00	48,33	8,67
45 58 48 59	38 64 61	52,50	54,33	– 1,83
45 58 48 61	38 64 59	53,00	53,67	– 0,67
45 58 59 61	38 64 48	55,75	50,00	5,75
45 64 48 59	38 58 61	54,00	52,33	1,67
45 64 48 61	38 58 59	54,50	51,67	2,83
45 64 59 61	38 58 48	57,25	48,00	9,25
45 48 59 61	38 58 64	53,25	53,33	– 0,08
58 64 48 59	38 45 61	57,25	48,00	9,25
58 64 48 61	38 45 59	57,75	47,33	10,42
58 64 59 61	38 45 48	60,50	43,67	16,83
58 48 59 61	38 45 64	56,50	49,00	7,50
64 48 59 61	38 45 58	58,00	47,00	11,00

El proceso que se muestra en Ejemplo 7.1.2 se denomina **test de aleatorización***. En un test de aleatorización se dividen aleatoriamente los datos observados en grupos para ver cómo es de probable que la diferencias observadas sean debidas únicamente al azar.

Nota: En la Sección 7.2 presentaremos un procedimiento conocido como el test t que a menudo proporciona una buena aproximación al test de aleatorización. El valor de 20/35 (0,57) calculado en el Ejemplo 7.1.2 se denomina P valor. (Ya hemos visto el uso de este término anteriormente, para la toma de decisiones en el contexto de la prueba de normalidad de Shapiro-Wilk en la Sección 4.4. El uso general de este término y de otros se explicará más en detalle en la Sección 7.2). Para los datos del Ejemplo 7.1.1, el test t produce un P valor de 0,54. Podemos ver este P valor del test t como una aproximación al P valor de 0,57 obtenido con el test de aleatorización.

Muestras más grandes

Cuando estamos tratando con muestras pequeñas, como en el Ejemplo 7.1.1, podemos hacer una lista con todos los posibles resultados de la asignación aleatoria de observaciones a grupos. El ejemplo siguiente muestra cómo manejar el caso de muestras grandes, donde no es posible hacer esa lista.

Ejemplo 7.1.3 Área de hojas

Un fisiólogo de plantas investigó el efecto de las tensiones mecánicas en el crecimiento de las plantas de soja. Un conjunto de plantas en semilleros separados se dividió en dos grupos. Las del primer grupo fueron sometidas a tensión agitándolas durante 20 minutos dos veces al día. Las del segundo grupo (el grupo de control) no fueron agitadas. Tras 16 días de crecimiento, las plantas fueron recolectadas y se midió el área total de las hojas (cm^2) de cada planta. Los datos se muestran en la Tabla 7.1.3 y se presentan en forma gráfica en la Figura 7.1.2².

Tabla 7.1.3	
Control	Tensión
314	283
320	312
310	291
340	259
299	216
268	201
345	267
271	326
285	241
media 305,8	266,2

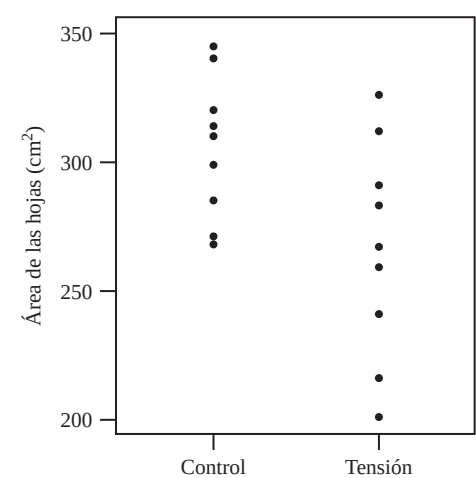


Figura 7.1.2 Diagramas de puntos en paralelo de las áreas de las hojas

La media de las plantas bajo tensión es menor que la de las plantas del grupo de control y la Figura 7.1.2 proporciona cierta evidencia visual de una diferencia entre los dos grupos. Por otra parte, los dos diagramas de puntos se solapan de forma considerable. Quizá el someter a tensión a las plantas, agitándolas, no tiene ningún efecto real en el área de las hojas y la diferencia tan elevada en este experimento ($305,8 - 266,2 = 39,6$) fue simplemente debida al azar. Es decir, podría ser que las condiciones «control» y «tensión» no tuvieran nada que ver con el área de las

* Muchas personas lo denominarían test de permutaciones, ya que requiere hacer una lista con todas las posibles permutaciones de los datos.

hojas. Si este es el caso, entonces los 18 semilleros pueden verse como de una única población, y la división en los grupos de «control» y «tensión» es arbitraria.

En el Ejemplo 7.1.2 pudimos hacer una lista con todas las posibles maneras en las que se podrían haber formado los dos grupos. Sin embargo, en el presente ejemplo hay 48.620 posibles formas de seleccionar 9 de los 18 semilleros como pertenecientes al grupo de control (y los otros 9 pertenecientes al grupo de tensión). Por tanto, no es posible crear una tabla similar a la Tabla 7.1.2 y hacer una lista con todas las posibilidades. Lo que podemos hacer, sin embargo, es muestrear aleatoriamente las 48.620 posibilidades. Una forma de hacer esto sería (1) escribir las 18 observaciones en 18 tarjetas; (2) barajar las tarjetas; (3) escoger aleatoriamente 9 de ellas como grupo de control y las otras 9 como grupo de tensión; (4) calcular la diferencia de las medias muestrales; (5) apuntar si el módulo de la diferencia de las medias muestrales es como mínimo 39,6; (6) repetir muchas veces los pasos (1)-(5).

Consideremos la fracción de veces que el módulo de la diferencia de las medias muestrales es al menos tan grande como el valor de 39,6 obtenido en el experimento. Esto es una medida de la evidencia contra la afirmación de que «la tensión que se crea sobre las plantas, agitándolas, no tiene un efecto real en el área de las hojas».

En vez de utilizar 18 tarjetas, podríamos utilizar una simulación por computador para hacer lo mismo. En una simulación con 1.000 pruebas hubo solo 36 en las que la diferencia de las medias muestrales fue mayor o igual que 39,6*. Esto indica que es poco probable que la diferencia observada de 39,6 sea debida al azar (el azar es solamente el 3,6 %), por lo que tenemos evidencia de que someter a tensión a las plantas produce un efecto. De hecho, parece que agitar los semilleros produjo una reducción en el área media de las hojas. ■

Nota: El procedimiento del test t (que será presentado en la Sección 7.2) produce un P valor de 0,033, es una buena aproximación al P valor de 0,036 obtenido mediante el test de aleatorización.

Ejercicios 7.1.1-7.1.3

7.1.1 Suponga que tenemos muestras de cinco hombres y cinco mujeres y hemos realizado un test de aleatorización para comparar entre sexos la variable Y = pulso. Suponga además que hemos encontrado que en 120 de los 252 posibles resultados bajo aleatorización la diferencia de las medias es al menos tan grande como la diferencia de las dos medias muestrales observadas. ¿Proporciona el test de aleatorización evidencia de que hay diferencias entre sexos con respecto al pulso? Justifique su respuesta utilizando los resultados de la aleatorización.

7.1.2 En una investigación de la posible influencia sobre los síntomas de la diabetes del cromo en la dieta, se alimentó algunas ratas con una dieta baja en cromo y otras con una dieta normal. Una variable de respuesta fue la actividad de la enzima del hígado GITH, que se midió utilizando una molécula marcada radiactivamente. La tabla que se presenta a continuación muestra los resultados, expresados como miles de unidades por minuto por gramo de hígado³. Las medias muestrales son 49,17 para la dieta baja en cromo y 51,90 para la dieta normal. Por tanto, la diferencia en las medias muestrales es de $-2,73$. Hay 10 posibles aleatorizaciones de las cinco observaciones en dos grupos, de tamaños tres y dos.

(a) Cree una lista de esas 10 aleatorizaciones (una de las cuales será la asignación original de observaciones a

grupos) y, para cada caso, calcule la media de la dieta baja en cromo menos la media de la dieta normal.

- (b) ¿Cuántas de las 10 aleatorizaciones producen una diferencia de las medias muestrales tan lejos de cero como $-2,73$ (la diferencia de las medias muestrales para nuestras muestras observadas)?
- (c) ¿Existe evidencia de que el cromo en la dieta afecta a la actividad de la enzima del hígado GITH? Justifique su respuesta utilizando los resultados de la aleatorización.

Dieta baja en cromo	Dieta normal
42,3	53,1
51,5	50,7
53,7	

7.1.3 La tabla siguiente muestra el número de colonias de bacterias presentes en varias placas de Petri, tras añadir bacterias *E. coli* a cada placa e incubarlas durante 24 horas. Las placas etiquetadas como «jabón» contenían una solución preparada con jabón ordinario. Las placas etiquetadas como «control» contenían una solución de agua estéril. (Estos datos son un subconjunto del conjunto de datos más

* En este caso, podríamos utilizar también un computador para considerar la diferencia de las medias en todas las 48.620 posibilidades, y anotar en cuántas de ellas las diferencias son mayores que 39,6 en módulo. Sin embargo, a medida que las muestras se hacen mayores, hacer una lista con todas las posibilidades puede ser computacionalmente costoso (incluso con computadores rápidos) y solo se obtendría un aumento marginal en exactitud con respecto a realizar las simulaciones como hemos descrito.

grande presentado en el Ejercicio 6.6.9). Las medias muestrales son 44 para el grupo de control y 39,7 para el grupo de jabón. Por tanto, la diferencia de las medias muestrales es 4,3, siendo mayor el grupo de control, como podría esperarse si el jabón fuera efectivo. Hay 20 posibles aleatorizaciones de las seis observaciones en dos grupos, cada uno de ellos de tamaño tres.

(a) Cree una lista de esas 20 aleatorizaciones (una de las cuales será la asignación original de observaciones a grupos) y, para cada caso, calcule la media del grupo de control menos la media del grupo de jabón.

(b) ¿Cuántas de las 20 aleatorizaciones producen una diferencia de las medidas muestrales al menos tan grande como 4,3?

(c) ¿Existe evidencia de que el jabón inhibe el crecimiento de *E. coli*? Justifique su respuesta utilizando los resultados de la aleatorización.

Control	Jabón
30	76
36	27
66	16

7.2 Contraste de hipótesis: el test t

En el Capítulo 6 vimos que se pueden comparar dos medias utilizando el intervalo de confianza para la diferencia ($\mu_1 - \mu_2$). Exploraremos ahora otro planteamiento para la comparación de medias: el procedimiento conocido como *contraste de hipótesis*. La idea general es formular como una hipótesis la afirmación de que μ_1 y μ_2 se diferencian y ver después si los datos proporcionan evidencia suficiente para dar soporte a esa hipótesis.

Las hipótesis nula y alternativa

La hipótesis de que μ_1 y μ_2 no son iguales se denomina **hipótesis alternativa** (o hipótesis de investigación), y se expresa abreviadamente como H_A . Se puede expresar así:

$$H_A: \mu_1 \neq \mu_2$$

Su antítesis es la **hipótesis nula**,

$$H_0: \mu_1 = \mu_2$$

que asevera que μ_1 y μ_2 son iguales. Un investigador expresaría en general esas hipótesis de manera más informal, como en el ejemplo siguiente.

Ejemplo 7.2.1 Tolueno y cerebro

El abuso de sustancias que contienen tolueno (por ejemplo, el pegamento) puede producir varios síntomas neurológicos. En una investigación sobre el mecanismo de estos efectos tóxicos, unos investigadores midieron las concentraciones de varios compuestos químicos en los cerebros de ratas que habían sido expuestas a una atmósfera cargada de tolueno, y también en ratas de control no expuestas a dicha atmósfera. Las concentraciones del compuesto químico del cerebro norepinefrina (NE) en la región medular del cerebro, para seis ratas expuestas al tolueno y cinco ratas de control, se muestran en la Tabla 7.2.1 y se presentan gráficamente en la Figura 7.2.1⁴.

La concentración media de NE en el grupo del tolueno ($\bar{y}_1 = 540,8$ ng/g) es sustancialmente mayor que la media del grupo de control ($\bar{y}_2 = 444,2$ ng/g). Podríamos preguntarnos si esta diferencia observada indica un fenómeno biológico real (el efecto del tolueno) o si la verdad podría ser que el tolueno no tiene efecto y la diferencia observada entre $\bar{y}_1$ y $\bar{y}_2$ refleja solo la variación por azar. Las hipótesis correspondientes, planteadas informalmente, serían:

H_0^* : El tolueno no tiene efecto en la concentración de NE en la médula de ratas.

H_A^* : El tolueno tiene algún efecto en la concentración de NE en la médula de ratas. ■

Indicaremos los planteamientos informales con símbolos diferentes (H_0^* y H_A^* en vez de H_0 y H_A) porque realizan aseveraciones diferentes. En el Ejemplo 7.2.1, la hipótesis alternativa informal realiza una afirmación muy fuerte (no solo existe una diferencia, sino que la diferencia está *causada* por el tolueno)*.

* Por supuesto, nuestras afirmaciones de H_0^* y H_A^* están abreviadas. Las afirmaciones completas deberían incluir todas las condiciones relevantes del experimento (ratas macho adultas, atmósfera de tolueno de 1.000 ppm durante 8 horas, etc.). Nuestro uso de afirmaciones averiadas no debería causar ninguna confusión.

Tabla 7.2.1 Concentración de NE (ng/g)		
	Tolueno (grupo 1)	Control (grupo 2)
	543	535
	523	385
	431	502
	635	412
	564	387
	549	
n	6	5
$\bar{y}$	540,8	444,2
s	66,1	69,6
ET	27	31

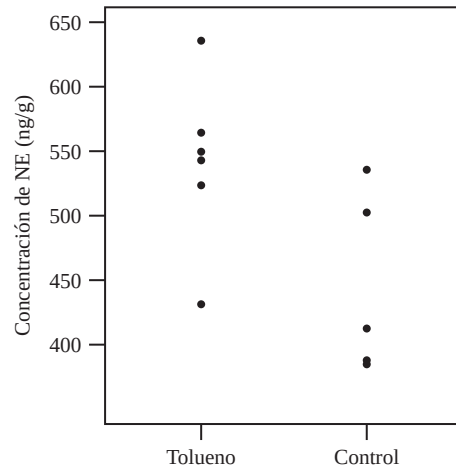


Figura 7.2.1 Diagramas de puntos en paralelo de la concentración de NE

Un **contraste de hipótesis estadístico** es un procedimiento para evaluar la fuerza de la evidencia presente en los datos para dar soporte a H_A . Se considera que los datos demuestran evidencia para H_A si las discrepancias con respecto a H_0 (la opuesta de H_A) no se pueden atribuir fácilmente al azar (es decir, al error de muestreo).

El estadístico t

Consideremos el problema de comprobar la hipótesis nula

$$H_0: \mu_1 = \mu_2$$

frente a la hipótesis alternativa

$$H_A: \mu_1 \neq \mu_2$$

Nótese que la hipótesis nula dice que las dos medias poblacionales son iguales, que es lo mismo que decir que la diferencia entre ellas es cero:

$$H_0: \mu_1 = \mu_2 \leftrightarrow H_0: \mu_1 - \mu_2 = 0$$

La hipótesis alternativa asevera que la diferencia no es cero:

$$H_A: \mu_1 \neq \mu_2 \leftrightarrow H_A: \mu_1 - \mu_2 \neq 0$$

El **test t** es un método estándar para escoger entre estas dos hipótesis. Para realizar el test t , el primer paso es calcular el **estadístico de contraste**, que para un test t se define como

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\text{ET}_{(\bar{y}_1 - \bar{y}_2)}}$$

Nótese que restamos cero de $\bar{y}_1 - \bar{y}_2$ porque la hipótesis H_0 plantea que $\mu_1 - \mu_2$ es igual a cero. Escribir « $(\bar{y}_1 - \bar{y}_2) - 0$ » nos recuerda lo que estamos comprobando. El subíndice «s» en t_s sirve como recordatorio de que este valor se calcula a partir de los datos. El valor t_s es el estadístico de contraste para el test t . Es decir, t_s proporciona el resumen de los datos que es la base para el procedimiento de contraste. Nótese la estructura de t_s : es una medida de lo lejos que está la diferencia de las medias muestrales (los valores $\bar{y}$) de la diferencia que esperaríamos ver si H_0 fuera cierta (diferencia cero), expresada en relación al ET de la diferencia (la cantidad de variación que esperaríamos ver en las diferencias de las medias de muestras aleatorias). Ilustraremos esta idea con un ejemplo.

Ejemplo 7.2.2 Tolueno y cerebro

Para los datos de NE en el cerebro del Ejemplo 7.2.1, el ET de $(\bar{Y}_1 - \bar{Y}_2)$ es

$$ET_{(\bar{Y}_1 - \bar{Y}_2)} = \sqrt{\frac{66,1^2}{6} + \frac{69,6^2}{5}} = 41,195$$

y el valor de t_s es

$$t_s = \frac{(540,8 - 444,2) - 0}{41,195} = 2,34$$

El estadístico t muestra que la diferencia entre $\bar{y}_1$ y $\bar{y}_2$ dista aproximadamente 2,3 ET de cero, la diferencia que esperaríamos ver si el tolueno no tuviera efectos sobre la NE. ■

¿Cómo podemos juzgar si nuestros datos presentan evidencia suficiente para H_A ? Una falta completa de evidencia (acuerdo *total* con H_0) se expresaría mediante medias muestrales que fueran idénticas y un estadístico t resultante igual a cero ($t_s = 0$). Pero, incluso si la hipótesis nula H_0 es cierta, no podemos esperar que t_s sea exactamente cero. Debemos esperar que las medidas muestrales difieran entre sí debido a la variabilidad del muestreo (medida mediante $ET_{(\bar{Y}_1 - \bar{Y}_2)}$). Afortunadamente, sabemos qué esperar con respecto a esta variabilidad en el muestreo. De hecho, no es probable que las diferencias por azar en los valores de $\bar{Y}$ superen un par de desviaciones típicas cuando la hipótesis nula es cierta. Para expresar esto de forma más precisa, se puede demostrar matemáticamente que

Si H_0 es cierta, entonces la distribución en el muestreo de t_s se aproxima bien mediante una distribución t de Student con grados de libertad dados por la fórmula (6.7.1)*.

La afirmación anterior es cierta si se cumplen ciertas condiciones. Brevemente: requerimos muestras aleatorias independientes de poblaciones distribuidas normalmente. Estas condiciones se considerarán en detalle en la Sección 7.9.

La esencia del procedimiento del test t es identificar si el valor observado t_s está en el interior de la distribución t de Student, como indica la Figura 7.2.2. Si t_s está cerca del centro, como en la Figura 7.2.2(a), entonces los datos se consideran compatibles con H_0 ya que la diferencia observada entre $(\bar{Y}_1 - \bar{Y}_2)$ se puede atribuir fácilmente a la variación por azar causada por el error de muestreo. (H_0 predice que las medidas muestrales serán iguales, ya que H_0 dice que las medias poblacionales son iguales).

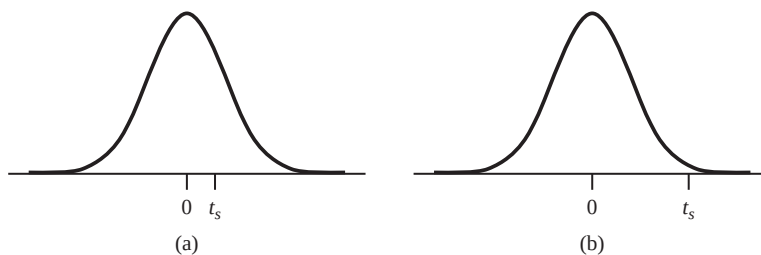


Figura 7.2.2 Esencia del test t . (a) Datos compatibles con H_0 (y por tanto falta de evidencia significativa para H_A); (b) datos incompatibles con H_0 (y por tanto evidencia significativa para H_A)

Si, por otra parte, t_s está lejos del centro, en la cola de la distribución t , como en la Figura 7.2.2(b), entonces los datos se consideran como evidencia a favor de H_A , ya que la desviación observada no se puede explicar fácilmente como debida a la variación por azar. Para expresar esto de otra forma, si H_0 es cierta, entonces es poco probable que t_s esté en las colas de la distribución, lejos del centro.

* Como indicamos en la Sección 6.8, una aproximación conservativa a la fórmula (6.7.1) es utilizar un número de grados de libertad dado por el mínimo de $n_1 - 1$ y $n_2 - 1$.

El P valor

Para juzgar si un valor observado t_s está «lejos» en la cola de la distribución t , necesitamos una vara de medir cuantitativa para localizar a t_s dentro de la distribución. Esta vara de medir la proporciona el P valor, que se puede definir (en el presente contexto) como sigue:

El **P valor** del contraste es el área bajo la curva t de Student en las colas dobles más allá de $-t_s$ y $+t_s$.

Por tanto, el P valor, que algunas veces se abrevia simplemente como « P », es el área sombreada en la Figura 7.2.3. Nótese que hemos definido el P valor como el área total en ambas colas. Esto se denomina algunas veces el P valor de «dos colas».

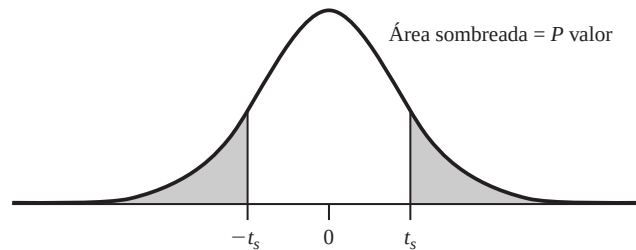


Figura 7.2.3 El P valor de dos colas del test t

Ejemplo 7.2.3 Tolueno y cerebro

Para los datos de NE en el cerebro del Ejemplo 7.2.1, el valor de t_s es 2,34. Podemos preguntarnos, «si H_0 fuera cierta, de forma que se pudiera esperar que $\bar{Y}_1 - \bar{Y}_2 = 0$ en promedio, ¿cuál es la probabilidad de que difiera de cero tanto como 2,34 veces ET?». El P valor responde a esta pregunta. La fórmula (6.7.1) da un resultado de 8,47 grados de libertad para estos datos. Por tanto, el P valor es el área bajo la curva t (con 8,47 grados de libertad) más allá de $\pm 2,34$. Esta área, que se puede obtener utilizando un computador, se muestra en la Figura 7.2.4 y vale 0,0454. ■

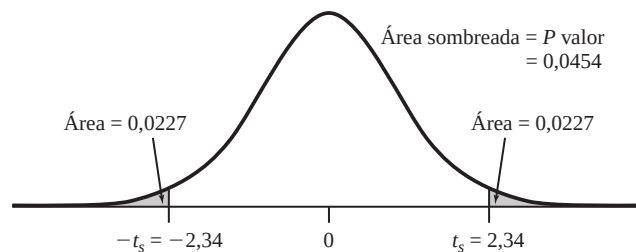


Figura 7.2.4 El P valor de dos colas para los datos del tolueno

Definición El P valor de un contraste de hipótesis es la probabilidad, calculada bajo la condición de que la hipótesis nula es verdadera, de que el estadístico de contraste sea al menos tan extremo como el valor del estadístico de contraste que ha sido obtenido realmente.

De la definición de P valor, se deduce que el **P valor es una medida de compatibilidad de los datos con H_0** y, por tanto, mide la **evidencia a favor de H_A** . Un P valor grande (cercano a 1) indica un valor de t_s cerca del centro de la distribución t (falta de evidencia a favor de H_A), mientras que un P valor pequeño (cercano a 0) indica un valor que está lejos en las colas de la distribución t (evidencia a favor de H_A).

Extracción de conclusiones de un test t

El P valor es una medida de la evidencia a favor de H_A en los datos, pero ¿dónde trazar la línea para determinar cuánta evidencia es suficiente? Muchos estarían de acuerdo en que un P valor = 0,0001 indica una evidencia muy fuerte mientras que un P valor = 0,80 indica una falta de evidencia, pero ¿qué sucede con los valores intermedios? Por ejemplo, debería un P valor = 0,10 ser considerado como una evidencia suficiente a favor de H_A ? La respuesta no es intuitivamente obvia.

En muchos trabajos científicos no es necesario trazar una frontera estricta. Sin embargo, en muchas situaciones debe tomarse una *decisión*. Por ejemplo, la Food and Drug Administration (FDA) debe decidir si los datos proporcionados por un fabricante farmacéutico son suficientes para aprobar justificadamente una medicación. Como otro ejemplo, un fabricante de fertilizantes debe decidir si la evidencia a favor de un nuevo fertilizante es suficiente para justificar el gasto de una investigación posterior.

Tomar una decisión requiere trazar una línea definida entre lo que es evidencia suficiente e insuficiente. El valor umbral, en la escala de P valores, se denomina **nivel de significación** del contraste y se representa mediante la letra griega α (alfa). El valor de α es elegido por cualquiera que tome la decisión. Valores comunes son $\alpha = 0,10, 0,05$ y $0,01$. Si el P valor de los datos es menor o igual que a α , se juzga que los datos proporcionan evidencia estadísticamente significativa a favor de H_A ; se dice también que H_0 **se rechaza**. Si el P valor de los datos es mayor que α , se dice que los datos no proporcionan evidencia suficiente para afirmar H_A y, por tanto, H_0 **no se rechaza**.

El siguiente ejemplo ilustra el uso del test t para tomar una decisión.

Ejemplo 7.2.4 Tolueno y cerebro

Para el experimento de NE en el cerebro del Ejemplo 7.2.1, los datos se resumen en la Tabla 7.2.2. Supongamos que elegimos tomar una decisión con un nivel de significación del 5 %, $\alpha = 0,05$. En el Ejemplo 7.2.3 obtuvimos que el P valor de estos datos es 0,0454. Esto significa que ha sucedido una de estas dos cosas: o bien (1) H_0 es cierta y tenemos un conjunto extraño de datos debido al azar o (2) H_0 es falsa. Si H_0 es verdadera, el tipo de discrepancia observada entre $\bar{y}_1$ e $\bar{y}_2$ ocurriría solo el 4,5 % de las veces. Debido a que el P valor, 0,0454, es menor que 0,05, rechazamos H_0 y concluimos que los datos proporcionan evidencia estadística suficiente a favor de H_A . La fuerza de la evidencia se expresa mediante la afirmación de que el P valor es 0,0454.

Tabla 7.2.2 Concentración de NE (ng/g)		
	Tolueno	Control
n	6	5
$\bar{y}$	540,8	444,2
s	66,1	69,6

Conclusión: Los datos proporcionan evidencia suficiente con un nivel de significación de 0,05 (P valor = 0,0454) de que el tolueno aumenta la concentración de NE*. ■

El siguiente ejemplo ilustra un test t en el que hay una falta de evidencia suficiente a favor de H_A con un nivel de significación de 0,05.

Ejemplo 7.2.5 Plantas rápidas

En el Ejemplo 6.7.1 vimos que la altura media de las plantas rápidas era más pequeña cuando se utilizaba ancy que cuando se utilizaba agua (el control). La Tabla 7.2.3 resume los datos. La diferencia entre las medias muestrales es $15,9 - 11,0 = 4,9$. El ET de la diferencia es

$$ET(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{4,8^2}{8} + \frac{4,7^2}{7}} = 2,46$$

* Debido a que la hipótesis alternativa era $H_A: \mu_1 \neq \mu_2$, algunos autores dirían, «concluimos que el tolueno afecta a la concentración de NE», en vez de decir que el tolueno aumenta la concentración de NE.

Tabla 7.2.3 Altura tras 14 días de plantas de control y de ancy		
	Control	Ancy
n	8	7
$\bar{y}$	15,9	11,0
s	4,8	4,7

Supongamos que escogemos utilizar $\alpha = 0,05$ en el contraste

$$H_0: \mu_1 = \mu_2 \text{ (p. ej., } \mu_1 - \mu_2 = 0)$$

frente a la hipótesis alternativa

$$H_A: \mu_1 \neq \mu_2 \text{ (p. ej., } \mu_1 - \mu_2 \neq 0)$$

El valor del estadístico de prueba es

$$t_s = \frac{(15,9 - 11,0) - 0}{2,46} = 1,99$$

La fórmula (6.7.1) da un valor de 12,8 grados de libertad para la distribución t . El P valor del contraste es la probabilidad de obtener un estadístico t que esté al menos tan alejado de cero como 1,99. La Figura 7.2.5 muestra que esta probabilidad es 0,0678. (Este P valor con 4 dígitos decimales se obtuvo utilizando un computador). Como el P valor es mayor que α , tenemos una evidencia suficiente a favor de H_A . Por tanto, no rechazamos H_0 . Es decir, estos datos no proporcionan evidencia suficiente para concluir que μ_1 y μ_2 son diferentes. La diferencia observada entre $\bar{y}_1$ y $\bar{y}_2$ podría haber sucedido fácilmente por azar.

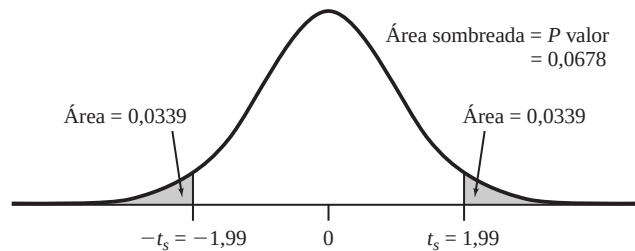


Figura 7.2.5 El P valor bilateral para los datos de ancy

Conclusión: Los datos *no* proporcionan evidencia suficiente (P valor = 0,0678) con un nivel de significación de 0,05 para concluir que el ancy y el agua tienen efectos diferentes en el crecimiento de las plantas rápidas (bajo las condiciones en las que se ha realizado el experimento). ■

Nótese cuidadosamente la forma de expresar la conclusión en el Ejemplo 7.2.5. *No* decimos que hay evidencia a favor de la hipótesis nula, sino solamente que hay insuficiente evidencia *en su contra*. Cuando no rechazamos H_0 , esto indica una falta de evidencia de que H_0 sea falsa, que no es lo mismo que una evidencia de que sea cierta. El astrónomo Carl Sagan (en otro contexto) resumió este principio sobre la evidencia en esta sucinta afirmación⁵:

La ausencia de evidencia no es evidencia de ausencia.

En otras palabras, el no rechazo de H_0 no es lo mismo que la *aceptación* de H_0 . (Para evitar confusión, es mejor no utilizar en absoluto la frase «aceptar H_0 »).

El no rechazo de H_0 indica que los datos son compatibles con H_0 , pero que dichos datos pueden *también* ser bastante compatibles con H_A . Por ejemplo, en el Ejemplo 7.2.5 obtuvimos que la diferencia observada entre las medias muestrales podría ser debida a la variación en el muestreo, pero este hecho no excluye la probabilidad de que la diferencia observada sea realmente debida a un efecto real causada por el ancy. (En la Sección 7.7 y en la Sección opcional 7.8 se presentan métodos para excluir posibles alternativas).

Al contrastar una hipótesis, el investigador empieza con el supuesto de que H_0 es cierta y se pregunta después si los datos contradicen dicho supuesto. Esta lógica puede tener sentido incluso si el investigador considera que la hipótesis nula es inverosímil. Por ejemplo, en el Ejemplo 7.2.5 se podría argumentar que casi ciertamente existe *alguna* diferencia (quizá muy pequeña) entre usar ancy y no usar ancy. El hecho de que no rechacemos H_0 no implica que aceptemos H_0 .

Uso de tablas frente a uso de tecnología

Al analizar datos, ¿cómo determinamos el P valor de un contraste? El *software* estadístico de los computadores y algunas calculadoras proporcionan valores exactos. Si esa tecnología no está disponible, entonces podemos utilizar la fórmula (6.7.1) para obtener los grados de libertad, pero redondeando hacia abajo para obtener un valor entero. Una alternativa conservativa a utilizar la fórmula (6.7.1) es utilizar como grados de libertad del contraste el valor mínimo de $n_1 - 1$ y $n_2 - 1$. Una alternativa por exceso es utilizar como grados de libertad del contraste $n_1 + n_2 - 2$. (La fórmula (6.7.1) siempre dará un valor de los grados de libertad entre el valor conservativo del mínimo de $n_1 - 1$ y $n_2 - 1$ y el valor por exceso de $n_1 + n_2 - 2$). Podemos basarnos en información limitada de la Tabla 4 para acotar el P valor, en vez de determinarlo exactamente. El P valor obtenido utilizando el planteamiento conservativo será algo mayor que el P valor exacto. El P valor obtenido utilizando la alternativa liberal será algo menor que el P valor exacto. El siguiente ejemplo ilustra el proceso de acotación.

Ejemplo 7.2.6 Plantas rápidas

Para los datos de crecimiento de plantas rápidas, el valor del estadístico t (tal como se determinó en el Ejemplo 7.2.5) es $t_s = 1,99$. El mínimo de $n_1 - 1$ y $n_2 - 1$ es $7 - 1 = 6$, por lo que un valor conservativo de los grados de libertad es 6. Un valor liberal de los grados de libertad es $8 + 7 - 2 = 13$. Se presenta a continuación una copia de parte de la Tabla 4, donde se han resaltado los números clave.

Probabilidad de la cola superior			
gl	0,05	0,04	0,03
6	1,943	2,104	2,313
7	1,895	2,046	2,241
8	1,860	2,004	2,189
9	1,833	1,973	2,150
10	1,812	1,948	2,120
11	1,796	1,928	2,096
12	1,782	1,912	2,076
13	1,771	1,899	2,060

Empezamos con el valor conservativo de los grados de libertad, 6. De la tabla anterior (o de la Tabla 4) obtenemos que $t_{6,0,05} = 1,943$ y $t_{6,0,04} = 2,104$. El correspondiente P valor conservativo, basado en una distribución t con 6 grados de libertad, está sombreado en la Figura 7.2.6. Como t_s está entre los valores críticos 0,04 y 0,05, el área de la cola superior debe estar entre 0,04 y 0,05. Por tanto, el P valor conservativo debe estar entre 0,08 y 0,10.

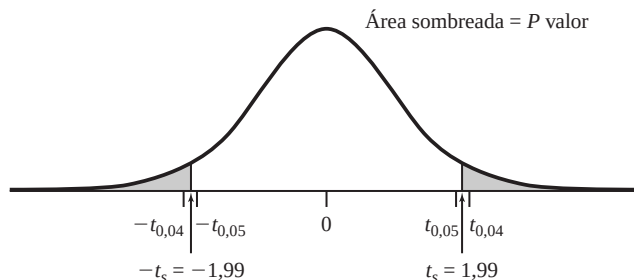


Figura 7.2.6 P valor conservativo para el Ejemplo 7.2.6

El valor por exceso de los grados de libertad es $8 + 7 - 2 = 13$. De la tabla anterior (o de la Tabla 4) obtenemos que $t_{13,0,04} = 1,899$ y $t_{13,0,03} = 2,060$. Como t_s está entre los valores críticos 0,03 y 0,04, el área de la cola superior debe estar entre 0,06 y 0,08. Por tanto, el P valor conservativo debe estar entre 0,06 y 0,08.

Reuniendo los dos casos, tenemos

$$0,06 < P \text{ valor} < 0,10$$

■

Si el valor de t_s no está entre los límites de la Tabla 4, entonces el P valor se encuentra acotado solo por un lado. Por ejemplo, si t_s es mayor que $t_{0,0005}$, entonces el P valor bilateral está acotado como

$$P \text{ valor} < 0,001$$

Presentación de los resultados de un test t

Al presentar los resultados de un test t , el investigador puede escoger entre tomar una decisión definida (afirmar que hay evidencia significativa a favor de H_A o que no hay evidencia significativa para dar soporte a H_A) con un nivel de significación especificado α , y escoger simplemente describir los resultados en frases como «hay una evidencia muy fuerte de que...» o «la evidencia sugiere que...» o «virtualmente la evidencia de que...». Al escribir un informe para su publicación, es muy deseable indicar el P valor de forma que el lector pueda tomar una decisión por sí mismo.

El término *significativo* se utiliza frecuentemente al presentar resultados. Por ejemplo, se dice que la diferencia observada es «estadísticamente significativa al nivel del 5 %» si es lo suficientemente grande como para justificar una evidencia significativa a favor de H_A con $\alpha = 0,05$. En el Ejemplo 7.2.4 vimos que la diferencia observada entre las dos medias muestrales de los datos del tolueno es estadísticamente significativa al nivel del 5 %, ya que el P valor es 0,0454, que es menor que 0,05. Por el contrario, los datos de las plantas rápidas del Ejemplo 7.2.5 no muestran una diferencia estadísticamente significativa al nivel del 5 %, ya que el P valor para esos datos es 0,0678. Sin embargo, la diferencia en las medias muestrales en los datos de las plantas rápidas es estadísticamente significativa al nivel de $\alpha = 0,10$, ya que el P valor es menor que 0,10. Cuando α no se especifica, se entiende en general que es 0,05. No obstante, hay que resaltar que α es un valor que se escoge arbitrariamente y que no existe nada «oficial» sobre que sea 0,05. Desafortunadamente, el término «significativo» se malinterpreta fácilmente y se debería utilizar con cuidado. Volveremos a este punto en la Sección 7.7.

Nota: En esta sección hemos considerado contrastes de la forma $H_0: \mu_1 = \mu_2$ (es decir, $\mu_1 - \mu_2 = 0$) frente a $H_A: \mu_1 \neq \mu_2$ (es decir, $\mu_1 - \mu_2 \neq 0$). Esta es la pareja de hipótesis más común; sin embargo, puede suceder que deseemos probar si μ_1 es mayor que μ_2 en alguna cantidad específica distinta de cero, por ejemplo c . Para probar $H_0: \mu_1 - \mu_2 = c$ frente a $H_A: \mu_1 - \mu_2 \neq c$ utilizaremos el test t con un estadístico de prueba dado por

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - c}{\text{ET}(\bar{y}_1 - \bar{y}_2)}$$

A partir de este punto, el contraste procede como antes (es decir, como en el caso de $c = 0$).

Ejercicios 7.2.1-7.2.17

[Nota: Las respuestas a las preguntas sobre contrastes de hipótesis deben incluir una afirmación de la conclusión en el contexto del ejercicio. (Véanse los Ejemplos 7.2.4 y 7.2.5)].

7.2.1 Para cada uno de los siguientes conjuntos de datos, utilice la Tabla 4 para acotar el P valor de dos colas de los datos analizados mediante el test t .

(a)

	Muestra 1	Muestra 2
n	4	3
$\bar{y}$	735	854
$\text{ET}(\bar{y}_1 - \bar{y}_2) = 38$ con $gl = 4$		

(b)

	Muestra 1	Muestra 2
n	7	7
$\bar{y}$	5,3	5,0
$\text{ET}(\bar{y}_1 - \bar{y}_2) = 0,24$ con $gl = 12$		

(c)

	Muestra 1	Muestra 2
n	15	20
$\bar{y}$	36	30
$\text{ET}(\bar{y}_1 - \bar{y}_2) = 1,3$ con $gl = 30$		

7.2.2 Para cada uno de los siguientes conjuntos de datos, utilice la Tabla 4 para acotar el P valor de dos colas de los datos analizados mediante el test t .

(a)

	Muestra 1	Muestra 2
n	8	5
$\bar{y}$	100,2	106,8
ET _($\bar{y}_1 - \bar{y}_2$) = 5,7 con gl = 10		

(b)

	Muestra 1	Muestra 2
n	8	8
$\bar{y}$	49,8	44,3
ET _($\bar{y}_1 - \bar{y}_2$) = 1,9 con gl = 13		

(c)

	Muestra 1	Muestra 2
n	10	15
$\bar{y}$	3,58	3,00
ET _($\bar{y}_1 - \bar{y}_2$) = 0,12 con gl = 19		

7.2.3 Para cada una de las siguientes situaciones, suponga que se contrasta la hipótesis $H_0: \mu_1 = \mu_2$ frente a $H_A: \mu_1 \neq \mu_2$. Indique si existe o no existe evidencia estadística a favor de H_A .

- (a) P valor = 0,085, $\alpha = 0,10$.
- (b) P valor = 0,065, $\alpha = 0,05$.
- (c) $t_s = 3,75$ con 19 grados de libertad, $\alpha = 0,01$.
- (d) $t_s = 1,85$ con 12 grados de libertad, $\alpha = 0,05$.

7.2.4 Para cada una de las siguientes situaciones, suponga que se contrasta la hipótesis $H_0: \mu_1 = \mu_2$ frente a $H_A: \mu_1 \neq \mu_2$. Indique si existe o no existe evidencia estadística a favor de H_A .

- (a) P valor = 0,046, $\alpha = 0,02$.
- (b) P valor = 0,033, $\alpha = 0,05$.
- (c) $t_s = 2,26$ con 5 grados de libertad, $\alpha = 0,10$.
- (d) $t_s = 1,94$ con 16 grados de libertad, $\alpha = 0,05$.

7.2.5 En un estudio sobre los requerimientos nutricionales del ganado, unos investigadores midieron las ganancias de peso de las vacas durante un periodo de 78 días. Para dos razas de vacas, Hereford (HH) y Brown Swiss/Hereford (SH), los resultados se muestran en la tabla siguiente⁶. [Nota: la fórmula (6.71) da un valor de 71,9 gl].

	HH	SH
n	33	51
$\bar{y}$	18,3	13,9
s	17,8	19,1

Utilice un test t para comparar las medias. Utilice $\alpha = 0,10$.

7.2.6 El espesor de grasa dorsal es una variable utilizada al evaluar la calidad de la carne de cerdo. Un científico de animales midió el espesor de la grasa dorsal (cm) en cerdos que habían seguido dos dietas diferentes, con los resultados dados en la siguiente tabla⁷.

	Dieta 1	Dieta 2
$\bar{y}$	3,49	3,05
s	0,40	0,40

Considere el uso del test t para comparar las dietas. Acote el P valor, suponiendo que el número de cerdos en cada dieta fue de

- (a) 5
- (b) 10
- (c) 15

Utilice $n_1 + n_2 - 2$ como aproximación al número de grados de libertad.

7.2.7 Los pacientes de enfermedad cardíaca a menudo sufren espasmos de las arterias coronarias. Como las aminos biológicas pueden tener un papel en estos espasmos, un equipo de investigación midió los niveles de amina en las arterias coronarias, obtenidos post mórtem de pacientes que habían muerto de enfermedad cardíaca y también de un grupo de control de pacientes que habían muerto por otras causas. La siguiente tabla resume la concentración de la amina serotonina⁸.

	Serotonina (ng/g)	
	Enfermedad cardíaca	Controles
n	8	12
$\bar{y}$	3.840	5.310
ET	850	640

- (a) Para estos datos, el ET de $(\bar{Y}_1 - \bar{Y}_2)$ es 1.064 y gl = 14,3 (que se puede redondear a 14). Utilice un test t para comparar las medias con un nivel de significación del 5 %.
- (b) Verifique el valor de ET_($\bar{y}_1 - \bar{y}_2$) dado en el apartado (a).

7.2.8 En un estudio sobre la cigarra periódica (*Magicalca septendecim*) unos investigadores midieron la longitud de la tibia de las patas traseras en las mudas de piel de 110 individuos. Los resultados de machos y hembras se muestran en la siguiente tabla⁹.

		Longitud de la tibia (μm)	
Grupo	n	Media	DT
Machos	60	78,42	2,87
Hembras	50	80,44	3,52

- (a) Utilice el test t para investigar la asociación de la longitud de la tibia con el género en esta especie. Utilice un nivel de significación del 5 %. [Nota: la fórmula (6.7.1) da un resultado de 94,3 gl].
- (b) Dados los datos anteriores, si se le diera la longitud de la tibia de un individuo de esta especie, ¿podría hacer una previsión con bastante confianza de su sexo? ¿Por qué o por qué no?
- (c) Repita el test t del apartado (a), suponiendo que las medias y desviaciones típicas son las que se dan en la tabla, pero que están basadas en solo la décima parte de los individuos (6 machos y 5 hembras). [Nota: la fórmula (6.7.1) da un resultado de 7,8 gl].

7.2.9 Se midió el flujo de sangre en el miocardio (FSM) en dos grupos de sujetos tras cinco minutos de ejercicio de bicicleta. El grupo de normoxia («oxígeno normal») recibió aire normal para respirar, mientras que al grupo de hipoxia se le proporcionó una mezcla de gases con oxígeno reducido, para simular condiciones de mucha altitud. Los resultados (ml/min/g) se muestran en la tabla siguiente¹⁰. [Nota: la fórmula (6.7.1) da un resultado de 12,2 gl].

	Normoxia	Hipoxia
	3,45	6,47
	3,09	5,69
	3,09	5,58
	2,65	5,27
	2,49	5,11
	2,33	4,88
	2,28	4,68
	2,24	3,50
	2,17	
	1,34	
n	10	8
$\bar{y}$	2,51	5,14
s	0,60	0,84

Utilice un test t para investigar el efecto de la hipoxia sobre el FSM. Utilice un valor de $\alpha = 0,05$.

7.2.10 En un estudio sobre el desarrollo de la glándula timo, unos investigadores pesaron las glándulas de 10 embriones de pollo. 5 de los embriones habían sido incubados durante 14 días y 5 habían sido incubados durante 15 días. Los pesos del timo se muestran en la tabla siguiente¹¹. [Nota: la fórmula (6.7.1) da un resultado de 7,7 gl].

- (a) Utilice un test t para comparar las medias con un valor de $\alpha = 0,10$.
- (b) Nótese que los embriones de pollo que fueron incubados más tiempo tuvieron un peso medio del timo menor.

¿Es sorprendente este «retroceso» o podría ser fácilmente atribuido al azar? Explique su respuesta.

	Peso del timo (mg)	
	14 días	15 días
	29,6	32,7
	21,5	40,3
	28,0	23,7
	34,6	25,2
	44,9	24,2
n	5	5
$\bar{y}$	37,72	29,22
s	8,73	7,19

7.2.11 Como parte de un experimento sobre el metabolismo de las raíces, un fisiólogo de plantas cultivó abedules en semilleros de un invernadero. Inundó cuatro semilleros con agua durante un día y utilizó otros cuatro como controles. Después recolectó los semilleros y analizó el contenido en ATP de las raíces. Los resultados (nmol de ATP por mg de tejido) se muestran en la tabla siguiente¹². [Nota: la fórmula (6.7.1) da un resultado de 5,6 gl].

	Inundados	Control
	1,45	1,70
	1,19	2,04
	1,05	1,49
	1,07	1,91
n	4	4
$\bar{y}$	1,190	1,785
s	0,184	0,241

Utilice un test t para investigar el efecto de la inundación. Utilice un valor de $\alpha = 0,10$.

7.2.12 Después de una operación, el volumen de sangre del paciente a menudo se reduce. En un estudio se midió el volumen total de plasma en circulación en cada paciente inmediatamente después de ser operado. Tras una transfusión de un «expansor de plasma» en el torrente sanguíneo, se midió el volumen de plasma de nuevo y se calculó el incremento de dicho volumen (ml). Dos de los expansores de plasma utilizados fueron albúmina (en 25 pacientes) y poligelatina (en 14 pacientes). La tabla que sigue muestra el incremento en el volumen de plasma¹³. [Nota: la fórmula (6.7.1) da un resultado de 33,6 gl].

Utilice un test t para comparar el incremento medio en el volumen de plasma después de los dos tratamientos. Utilice un valor de $\alpha = 0,10$.

	Albúmina	Poligelatina
n	25	14
incremento medio	490	240
ET	60	30

7.2.13 Unos investigadores en nutrición realizaron una investigación sobre dos dietas altas en fibra cuyo objetivo era reducir el nivel de colesterol en suero. Se asignaron aleatoriamente 20 hombres con niveles altos de colesterol en suero para recibir una dieta «avena» o una dieta «judía» durante 21 días. La tabla que sigue resume el descenso (después menos antes) en los niveles de colesterol en suero¹⁴. Utilice un test t para comparar las dietas con un nivel de significación del 5 %. [Nota: la fórmula (6.7.1) da un resultado de 17,9 gl].

	Descenso del colesterol (mg/dl)		
Dieta	n	Media	DT
Avena	10	53,6	31,1
Judía	10	55,5	29,4

7.2.14 Suponga que se ha realizado un test t , con $\alpha = 0,05$ y un P valor de 0,03. Indique si cada una de las siguientes afirmaciones es verdadera o falsa y explique por qué.

- Rechazamos H_0 con $\alpha = 0,05$.
- Tenemos evidencia significativa a favor de H_A con $\alpha = 0,05$.
- Rechazaríamos H_0 si α fuera 0,10.
- No tenemos evidencia significativa a favor de H_A con $\alpha = 0,10$.
- Si H_0 es cierta, la probabilidad de obtener un estadístico de contraste al menos tan extremo como el valor de t_s obtenido realmente es del 3 %.
- Hay un 3 % de probabilidad de que H_0 sea cierta.

7.2.15 Suponga que se ha realizado un test t , con $\alpha = 0,10$ y un P valor de 0,07. Indique si cada una de las siguientes afirmaciones es verdadera o falsa y explique por qué.

- Rechazamos H_0 con $\alpha = 0,10$.
- Tenemos evidencia significativa a favor de H_A con $\alpha = 0,10$.
- Rechazaríamos H_0 si α fuera 0,05.
- No tenemos evidencia significativa a favor de H_A con $\alpha = 0,10$.
- La probabilidad de que $\bar{Y}_1$ sea mayor que $\bar{Y}_2$ es 0,07.

7.2.16 La tabla siguiente muestra el número de colonias de bacterias presentes en varias placas de Petri, después de añadir *E. coli* a dichas placas e incubarlas durante 24 horas. Las placas «jabón» contenían una solución preparada con jabón ordinario. Las placas «control» contenían una solución de agua estéril. (Estos datos se presentaron en el Ejercicio 6.6.9).

Utilice un test t para investigar si el jabón afecta al número de colonias de bacterias que se forman. Utilice $\alpha = 0,10$. [Nota: para estos datos, la fórmula (6.7.1) da un resultado de 10,4 gl].

	Control	Jabón
	30	76
	36	27
	66	16
	21	30
	63	26
	38	46
	35	6
	45	
n	8	7
$\bar{y}$	41,8	32,4
s	15,6	22,8
ET	5,5	8,6

7.2.17 Unos investigadores estudiaron el efecto de un fertilizante de plantas domésticas en el crecimiento de los brotes de rábanos. Seleccionaron aleatoriamente algunas semillas de rábanos para que sirvieran de control, y se plantaron otras en macetas de aluminio a las que añadieron barras de fertilizante. Las otras condiciones se mantuvieron iguales entre los dos grupos. La tabla siguiente muestra los datos de las alturas de las plantas (en cm) dos semanas después de la germinación¹⁵.

Utilice un test t para investigar si el fertilizante tuvo algún efecto en el crecimiento medio de los brotes de rábanos. Utilice $\alpha = 0,10$. [Nota: para estos datos, la fórmula (6.7.1) da un resultado de 53,5 gl].

Control		Fertilizante	
3,4	1,6	2,8	1,9
4,4	2,9	1,9	2,7
3,5	2,3	3,6	2,3
2,9	2,8	1,2	1,8
2,7	2,5	2,4	2,7
2,6	2,3	2,2	2,6
3,7	1,6	3,6	1,3
2,7	1,6	1,2	3,0
2,3	3,0	0,9	1,4
2,0	2,3	1,5	1,2
1,8	3,2	2,4	2,6
2,3	2,0	1,7	1,8
2,4	2,6	1,4	1,7
2,5	2,4	1,8	1,5
n	28	28	
$\bar{y}$	2,58	2,04	
s	0,65	0,72	

7.3 Comentarios adicionales sobre el test t

En esta sección presentaremos con más detalle el método y la interpretación del test t .

Relación entre el test y el intervalo de confianza

Hay una estrecha conexión entre el planteamiento del intervalo de confianza y el planteamiento del contraste de hipótesis para comparar μ_1 y μ_2 . Consideremos, por ejemplo, un intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ y su relación con el test t con un nivel de significación del 5 %. El test t y el intervalo de confianza utilizan las mismas tres cantidades: $(\bar{Y}_1 - \bar{Y}_2)$, $ET_{(\bar{Y}_1 - \bar{Y}_2)}$ y $t_{0,025}$, pero las utilizan de forma diferente.

En el test t , cuando $\alpha = 0,05$, tenemos evidencia significativa a favor de H_A (y por tanto rechazamos H_0) si el P valor es menor o igual que 0,05. Esto ocurre si y solo si el estadístico de contraste, t_s , está en la cola de la distribución t , en o más allá de $\pm t_{0,025}$. Si el módulo de t_s (simbolizado como $|t_s|$) es mayor o igual que $t_{0,025}$, entonces el P valor es menor o igual que 0,05 y tenemos evidencia significativa a favor de H_A . Si $|t_s|$ es menor que $t_{0,025}$, entonces el P valor es mayor que 0,05 y no tenemos evidencia significativa a favor de H_A . La Figura 7.3.1 muestra esta relación.

Por tanto, hay una falta de evidencia significativa a favor de $H_A: \mu_1 - \mu_2 \neq 0$ si y solo si $|t_s| < t_{0,025}$. Es decir, hay falta de evidencia significativa a favor de H_A cuando

$$\frac{|\bar{y}_1 - \bar{y}_2|}{ET_{(\bar{Y}_1 - \bar{Y}_2)}} < t_{0,025}$$

Esto es equivalente a

$$|\bar{y}_1 - \bar{y}_2| < t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)}$$

o

$$-t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)} < (\bar{y}_1 - \bar{y}_2) < t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)}$$

que es equivalente a

$$-(\bar{y}_1 - \bar{y}_2) - t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)} < 0 < -(\bar{y}_1 - \bar{y}_2) + t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)}$$

o

$$(\bar{y}_1 - \bar{y}_2) + t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)} > 0 > (\bar{y}_1 - \bar{y}_2) - t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)}$$

o

$$(\bar{y}_1 - \bar{y}_2) - t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)} < 0 < (\bar{y}_1 - \bar{y}_2) + t_{0,025} ET_{(\bar{Y}_1 - \bar{Y}_2)}$$

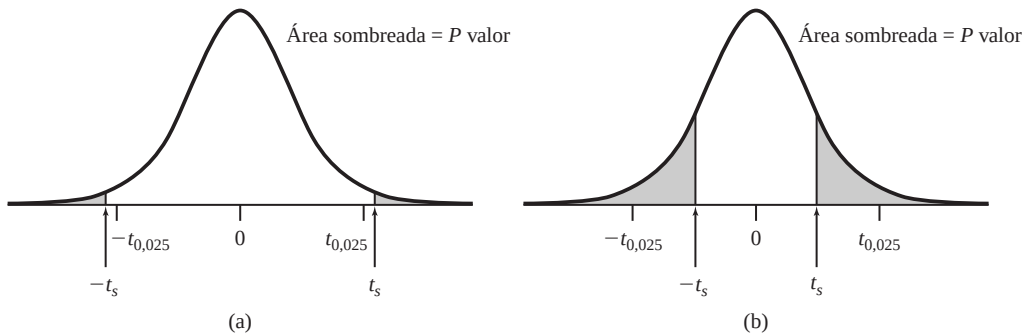


Figura 7.3.1 Posibles resultados del test t con $\alpha = 0,05$. (a) Si $|t_s| \geq t_{0,025}$, entonces $P \text{ valor} \leq 0,05$ y hay evidencia significativa a favor de H_A (por tanto se rechaza H_0). (b) Si $|t_s| < t_{0,025}$, entonces $P \text{ valor} > 0,05$ y no hay evidencia significativa a favor de H_A .

Por tanto, hemos demostrado que hay falta de evidencia significativa a favor de $H_A: \mu_1 - \mu_2 \neq 0$ si y solo si el intervalo de confianza para $(\mu_1 - \mu_2)$ incluye al cero. Por el contrario, si el intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ no

incluye al cero, entonces tenemos evidencia significativa a favor de $H_A: \mu_1 - \mu_2 \neq 0$ con $\alpha = 0,05$. (Se mantiene la misma relación entre el intervalo de confianza del 90 % y el test con $\alpha = 0,10$, y así sucesivamente). Ilustraremos esta idea con un ejemplo.

Ejemplo 7.3.1 Longitudes de cangrejos de río

Unos biólogos tomaron muestras de la especie de cangrejos de río *Orconectes sanborii* de los ríos del centro de Ohio, el río Upper Cuyahoga (CUY) y el arroyo East Fork of Pine (EFP), y midieron la longitud (mm) de cada cangrejo de río capturado¹⁶. La Tabla 7.3.1 muestra los estadísticos resumen. La Figura 7.3.2 muestra diagramas de caja en paralelo de los datos. La distribución de la muestra de EFP está desplazada hacia abajo con respecto a la distribución de la muestra de CUY. Ambas distribuciones son razonablemente simétricas.

Tabla 7.3.1 Datos de cangrejos de río: longitud (mm)		
	CUY	EFP
n	30	30
$\bar{y}$	22,91	21,97
s	3,78	2,90

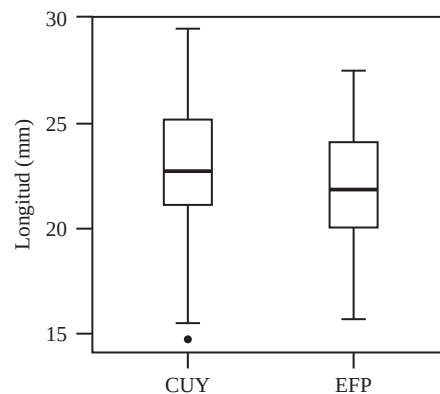


Figura 7.3.2 Diagramas de caja de los datos de los cangrejos de río

Para estos datos, los dos ET son $3,78/\sqrt{30} = 0,69$ y $2,90/\sqrt{30} = 0,53$ para CUY y EFP, respectivamente. Los grados de libertad son

$$gl = \frac{(0,69^2 + 0,53^2)^2}{0,69^4/30 + 0,53^4/30} = 56,3$$

Las cantidades necesarias para realizar el test t con $\alpha = 0,05$ son

$$\bar{y}_1 - \bar{y}_2 = 22,91 - 21,97 = 0,94$$

y

$$ET_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{0,69^2 + 0,53^2} = 0,87$$

El estadístico de contraste es

$$t_s = \frac{(22,91 - 21,97) - 0}{0,87} = \frac{0,94}{0,87} = 1,08$$

El P valor en este caso (obtenido utilizando un computador) es 0,2850, es mayor que 0,05, por lo que no rechazamos H_0 . (Una mirada rápida a la Tabla 4, utilizando $gl = 50$, muestra que el P valor está entre 0,20 y 0,40).

Si construimos un intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ obtenemos

$$0,94 \pm 2,006 \times 0,87$$

o $(-2,68, 0,81)^*$.

* El valor de $t_{0,025} = 2,006$ se basa en 56,3 grados de libertad. Si utilizáramos 50 grados de libertad (es decir, si hubiéramos utilizado la Tabla 4, en vez de un computador, el multiplicador t sería 2,009. Esto casi no causa diferencia en el intervalo de confianza resultante.

El intervalo de confianza incluye al cero, que es consistente con el hecho de que no hay evidencia significativa a favor de $H_A: \mu_1 - \mu_2 \neq 0$ en el test t . Nótese que esta equivalencia entre el test y el intervalo de confianza es de sentido común. De acuerdo con el intervalo de confianza, μ_1 puede ser como mucho inferior en 2,68, o como mucho superior en 0,81, a μ_2 . Es natural entonces decir que no estamos seguros de si es mayor que (o menor que, o igual a) μ_2 . ■

En el contexto del método de la t de Student, el planteamiento del intervalo de confianza y el planteamiento del contraste de hipótesis son formas diferentes de utilizar la misma información básica. El intervalo de confianza tiene la ventaja de que indica la magnitud de la diferencia entre μ_1 y μ_2 . El planteamiento del contraste tiene la ventaja de que el P valor describe en una escala continua de la fuerza de la evidencia de que μ_1 y μ_2 son realmente diferentes. En la Sección 7.7 continuaremos explorando el uso de un intervalo de confianza para complementar la interpretación de un test t . En capítulos posteriores encontraremos otro contraste de hipótesis que no se pueden complementar tan fácilmente con un intervalo de confianza.

Interpretación de α

Al analizar datos o tomar una decisión basada en dichos datos, a menudo necesitaremos escoger en nivel de significación α . ¿Cómo podemos saber si escoger $\alpha = 0,05$ o $\alpha = 0,01$ o algún otro valor? Para hacer este juicio, es útil tener una interpretación operacional de α . Daremos ahora esta interpretación.

Recuérdese de la Sección 7.2 que la distribución muestral de t_s , si H_0 es cierta, es una distribución t de Student. Supongamos ahora, por concretar, que $gl = 60$ y que se escoge α igual a 0,05. El valor crítico (de la Tabla 4) es $t_{0,025} = 2,000$. La Figura 7.3.3 muestra la distribución t de Student y los valores $\pm 2,000$. El área total sombreada en la figura es 0,05. Está dividida en dos partes iguales, cada una de ellas de área igual a 0,025. Podemos ver la Figura 7.3.3 como una vía formal para decidir si la evidencia es suficientemente fuerte para dar un soporte significativo a H_A . Si el valor observado de t_s está en las regiones marcadas del eje t_s , entonces hay evidencia significativa a favor de H_A . Pero la probabilidad de que esto suceda si H_0 es cierta es del 5 %. Por tanto, podemos decir que

$$\Pr\{\text{los datos proporcionen evidencia significativa a favor de } H_A\} = 0,05 \text{ si } H_0 \text{ es cierta}$$

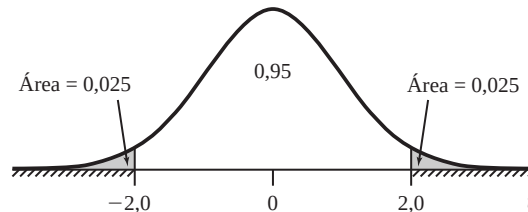


Figura 7.3.3 Un test t con $\alpha = 0,05$. Hay evidencia significativa a favor de H_A si t_s está en la región marcada

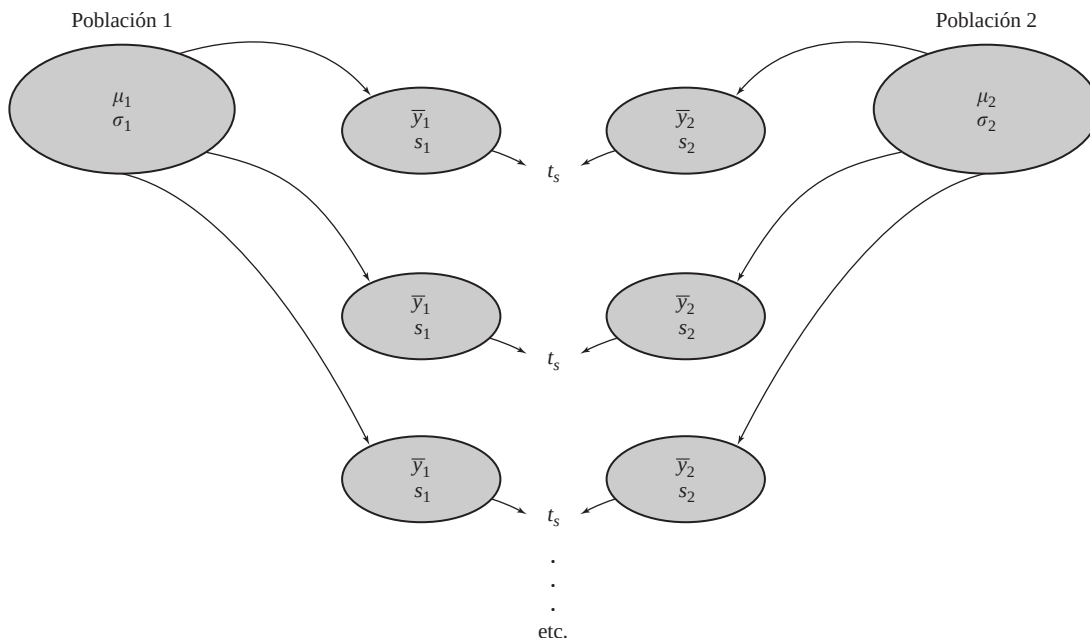
Esta probabilidad tiene sentido en el contexto de un metaestudio (que se muestra en la Figura 7.3.4), en el que muestreemos repetidamente dos poblaciones y calculamos un valor de t_s . Es importante darse cuenta de que la probabilidad se refiere a una situación en la que H_0 es cierta. Para representar concretamente esta situación, invitamos al lector a suspender la incredulidad por un momento y embarcarse en un viaje imaginario en el Ejemplo 7.3.2.

Ejemplo 7.3.2 Música y caléndulas*

Imagine que la comunidad científica ha desarrollado un gran interés en la influencia de la música sobre el crecimiento de las caléndulas. Una escuela de investigación se centra en si la música escrita por Bach o Mozart produce plantas más altas. Las plantas se asignan aleatoriamente a música de Bach (tratamiento 1) o música de Mozart (tratamiento 2) y, tras un periodo adecuado de escucha, se recogen datos sobre la altura de las plantas. La hipótesis nula es

$$H_0: \text{Las caléndulas responden igualmente bien a Bach o a Mozart}$$

* Este ejemplo es intencionadamente extravagante.

Figura 7.3.4 Metaestudio para el test t

o

$$H_0: \mu_1 = \mu_2$$

siendo

μ_1 = Altura media de las caléndulas expuestas a Bach
 μ_2 = Altura media de las caléndulas expuestas a Mozart

Supongamos a efectos del argumento que H_0 es de hecho cierta. Imaginemos ahora que muchos investigadores realizan el experimento de Bach frente a Mozart, y que cada experimento produce datos con 60 grados de libertad. Supongamos que cada investigador analiza sus datos mediante un test t con $\alpha = 0,05$. ¿Qué conclusiones alcanzarán los investigadores? En el metaestudio de la Figura 7.3.4, supongamos que cada pareja de muestras representa a un investigador diferente. Como estamos suponiendo que μ_1 y μ_2 son realmente iguales, los valores de t_s se desviarán de cero solo debido al error de muestreo causado por el azar. Si todos los investigadores se juntaran y realizaran una distribución de frecuencias de sus valores de t_s , dicha distribución seguiría una curva t de Student con 60 grados de libertad. Los investigadores tomarían sus decisiones como se indica en la Figura 7.3.3, por lo que podríamos esperar que tuvieran las siguientes experiencias:

- El 95 % de ellos no encontrarían (correctamente) evidencia significativa a favor de H_A .
- El 2,5 % de ellos encontrarían evidencia significativa a favor de H_A y concluirían (incorrectamente) que las plantas prefieren a Bach.
- El 2,5 % de ellos encontrarían evidencia significativa a favor de H_A y concluirían (incorrectamente) que las plantas prefieren a Mozart.

Por tanto, un total del 5 % de los investigadores encontrarían evidencia significativa a favor de la hipótesis alternativa. ■

El Ejemplo 7.3.2 proporciona una forma de interpretar α . Por supuesto, al analizar datos, no estamos haciendo un metaestudio, sino un único experimento. Cuando realizamos un test t con un nivel de significación del 5 %, estamos tomando el papel de uno de los investigadores del Ejemplo 7.3.2, y los otros son imaginarios. Si obtenemos evidencia significativa a favor de H_A , hay dos posibilidades:

1. H_A es de hecho cierta.
2. H_0 es de hecho cierta, pero somos parte del 5 % desafortunado que obtuvo datos que proporcionaron evidencia significativa a favor de H_A . En este caso, podemos pensar que la evidencia significativa a favor de H_A «dispara una falsa alarma».

Nos sentimos «confiados» al afirmar que nuestra evidencia a favor de H_A es significativa porque la segunda posibilidad es poco probable (suponiendo que consideramos que el 5 % es un porcentaje pequeño). Por supuesto, nunca sabremos (a menos que alguien repita el experimento) si somos o no somos parte del desafortunado 5 %.

Nivel de significación frente a P valor A veces los estudiantes encuentran difícil distinguir entre nivel de significación (α) y P valor*. En el test t , tanto α como el P valor son áreas en las colas bajo la curva t de Student. Pero α es un valor arbitrario especificado previamente. Puede ser (y debería ser) elegido antes de ver los datos. Por el contrario, el P valor está determinado por los datos. De hecho, dar el P valor es una forma de describir los datos. Puede resultar útil en este punto comparar la Figura 7.2.3 con la Figura 7.3.3. El área sombreada representa el P valor en la primera figura y α en la segunda figura.

Errores Tipo I y Tipo II

Hemos visto que α se puede interpretar como una probabilidad:

$$\alpha = \Pr\{\text{obtener evidencia significativa a favor de } H_A\} \text{ si } H_0 \text{ es cierta}$$

Afirmar que los datos proporcionan evidencia que de forma significativa da soporte a H_A cuando H_0 es cierta se denomina **error de Tipo I**. Al escoger α , estamos eligiendo nuestro nivel de protección contra el error de Tipo I. Muchos investigadores consideran el 5 % como un riesgo aceptablemente pequeño. Si no consideramos que el 5 % sea suficientemente pequeño, podemos escoger utilizar un valor más conservativo de α como $\alpha = 0,01$. En este caso, el porcentaje de hipótesis nulas verdaderas que rechazaríamos no sería del 5 %, sino del 1 %.

En la práctica, la selección de α puede depender del contexto del experimento concreto. Por ejemplo, una agencia de regulación podría exigir una prueba más exacta de la eficacia en el caso de una medicina tóxica que en el caso de una relativamente inocua. También, la elección personal de α puede estar influida por su opinión previa sobre el fenómeno bajo estudio. Por ejemplo, supongamos que un agrónomo es escéptico sobre las afirmaciones de un cierto tratamiento del suelo. Al evaluar un nuevo estudio del tratamiento, podía expresar su escepticismo escogiendo un nivel de significación muy conservador (por ejemplo, $\alpha = 0,001$), indicando así que se necesitaría mucha evidencia para convencerle de que el tratamiento es efectivo. Por esta razón, los informes escritos de una investigación deben incluir un P valor, de forma que cada lector sea libre de escoger su propio valor de α al evaluar los resultados del informe.

Si H_A es cierta, pero no observamos evidencia suficiente a favor de H_A , entonces hemos cometido un **error de Tipo II**. La Tabla 7.3.2 muestra las situaciones en las que pueden ocurrir errores de Tipo I y de Tipo II. Por ejemplo, si encontramos evidencia significativa a favor de H_A , entonces eliminamos la posibilidad de cometer un error de Tipo II, pero al rechazar H_0 podemos cometer un error de Tipo I.

Tabla 7.3.2 Posibles resultados al contrastar H_0			
		Situación real	
		H_0 verdadera	H_A verdadera
NUESTRA DECISIÓN	Falta de evidencia significativa a favor de H_A	Correcto	Error de Tipo II
	Evidencia significativa a favor de H_A	Error de Tipo I	Correcto

* Desafortunadamente el término «nivel de significación» no se utiliza consistentemente por toda la gente que escribe sobre la estadística. Unos pocos autores utilizan los términos «nivel de significación» o «probabilidad de significación» donde nosotros hemos utilizado « P valor».

Las consecuencias de los errores de Tipo I y de Tipo II pueden ser muy diferentes. Los dos ejemplos siguientes muestran algo de la variedad de estas consecuencias.

Ejemplo 7.3.3 Marihuana y glándula pituitaria

Los cannabinoides, que son sustancias contenidas en la marihuana, se pueden transmitir de madre a hijo a través de la placenta y a través de la leche. Supongamos que realizamos el siguiente experimento en ratones hembra embarazadas: suministramos a un grupo de ratones una dosis de cannabinoides y mantenemos otro grupo como control. Evaluamos seguidamente la función de la glándula pituitaria de la descendencia. Las hipótesis serían

H_0 : Los cannabinoides no afectan a la glándula pituitaria de la descendencia.

H_A : Los cannabinoides afectan a la glándula pituitaria de la descendencia.

Si de hecho los cannabinoides no afectan a la glándula pituitaria de la descendencia, pero concluimos que nuestros datos proporcionan evidencia significativa a favor de H_A , estamos cometiendo un error de Tipo I. La consecuencia podría ser una alarma innecesaria si la conclusión se hiciera pública. Por otra parte, si los cannabinoides afectaran a la glándula pituitaria de la descendencia, pero el resultado de nuestro test t es una falta de evidencia significativa a favor de H_A , estaríamos cometiendo un error de Tipo II. Una consecuencia podría ser una autocomplacencia injustificable por parte de las madres que fuman marihuana. ■

Ejemplo 7.3.4 Inmunoterapia

La quimioterapia es el tratamiento estándar de un cierto cáncer. Supongamos que realizamos un estudio clínico para probar la eficacia de suplementar a la quimioterapia con inmunoterapia (estimulación del sistema inmune). Los pacientes reciben o bien solo quimioterapia o quimioterapia más inmunoterapia. Las hipótesis serían

H_0 : La inmunoterapia no es efectiva para mejorar la supervivencia.

H_A : La inmunoterapia afecta la supervivencia.

Si la inmunoterapia realmente no es efectiva, pero concluimos que nuestros datos proporcionan evidencia significativa a favor de H_A y concluimos por tanto que la inmunoterapia es efectiva, entonces hemos cometido un error de Tipo I. La consecuencia, si esta conclusión es aceptada por la comunidad médica, podría ser un uso generalizado de inmunoterapia desagradable, peligrosa e inefectiva. Si, por otra parte, la inmunoterapia es realmente efectiva pero nuestros datos no nos permiten detectar ese hecho (quizá porque nuestros tamaños de muestra son demasiado pequeños), entonces hemos cometido un error de Tipo II, con consecuencias bastante diferentes de las de un error de Tipo I. Se continuaría utilizando el tratamiento estándar hasta que alguien proporcionara una evidencia convincente de que la inmunoterapia suplementaria es efectiva. Si todavía «creemos» en la inmunoterapia, podríamos realizar otro estudio (quizá con muestras más grandes) para intentar de nuevo establecer su efectividad. ■

Como ilustran los ejemplos anteriores, las consecuencias de un error de Tipo I son generalmente bastante diferentes de las de un error de Tipo II. Las probabilidades que los dos tipos de error pueden ser muy diferentes también. El nivel de significación α es la probabilidad de obtener evidencia significativa a favor de H_A si H_0 es cierta. Como α se escoge a voluntad, el procedimiento de contraste de hipótesis nos «protege» contra el error de Tipo I proporcionándonos control sobre el riesgo de ese error. Este control es independiente del tamaño de la muestra y de otros factores. Por el contrario, la probabilidad de un error de Tipo II depende de muchos factores, y puede ser grande o pequeña. En particular, un experimento con tamaños de muestra pequeños a menudo tendrá un alto riesgo de errores de Tipo II.

Estamos ahora en posición de volver a examinar el aforismo de Carl Sagan de que «la ausencia de evidencia no es evidencia de ausencia». Como el riesgo de error de Tipo I está controlado y el de el error de Tipo II no lo está, nuestro estado de conocimiento es mucho más fuerte tras el rechazo de una hipótesis nula que tras su no rechazo. Por ejemplo, supongamos que estamos probando si un cierto aditivo del suelo es efectivo en incrementar la cosecha de maíz. Si encontramos evidencia significativa a favor de H_A y afirmamos que el aditivo es efectivo, entonces o bien (1) estamos en lo cierto o (2) hemos cometido un error de Tipo I. Como el riesgo de un error de Tipo I está controlado, podemos estar relativamente confiados en nuestra conclusión de que el aditivo es efectivo (aunque no necesariamente muy efectivo). Supongamos, por otra parte, que los datos son tales que hay una falta de evidencia sobre la efectividad del aditivo, es decir, no tenemos evidencia a favor de H_A . Entonces, o bien (1) estamos en lo cierto (es decir, H_0 es cierta), o (2) hemos

cometido un error de Tipo II. Como el riesgo de cometer un error de Tipo II puede ser bastante alto, no podemos afirmar fiablemente que el aditivo es inefectivo. Para justificar la afirmación de que el aditivo es inefectivo, necesitaríamos suplementar nuestro contraste de hipótesis con un análisis posterior, como un intervalo de confianza o un análisis de la probabilidad de cometer un error de Tipo II. Consideraremos esto con más detalle en las Secciones 7.6 y 7.7.

Potencia

Como hemos visto, el error de Tipo II es un concepto importante. La probabilidad de cometer un error de Tipo II se denomina β :

$$\beta = \Pr\{\text{falta de evidencia significativa a favor de } H_A\} \text{ cuando } H_A \text{ es cierta}$$

La probabilidad de no cometer un error de Tipo II cuando H_A es cierta (es decir, la probabilidad de tener evidencia significativa a favor de H_A cuando H_A es cierta) se denomina potencia de un contraste estadístico:

$$\text{Potencia} = 1 - \beta = \Pr\{\text{evidencia significativa a favor de } H_A\} \text{ cuando } H_A \text{ es cierta}$$

Por tanto, la potencia de un test t es una medida de la sensibilidad de dicho test, o la capacidad del procedimiento del test para detectar una diferencia entre μ_1 y μ_2 cuando realmente esa diferencia *sí* existe. En este sentido la potencia es análoga a la potencia de resolución de un microscopio.

La potencia de un contraste estadístico depende de muchos factores en una investigación, incluyendo los tamaños de la muestra, la variabilidad inherente de las observaciones y la magnitud de la diferencia entre μ_1 y μ_2 . Si todo lo demás es igual, utilizar muestras más grandes dará más información y, por tanto, aumentará la potencia. Además, veremos que algunos contrastes estadísticos pueden ser más potentes que otros y que algunos diseños de estudios pueden ser más potentes que otros.

La planificación de una investigación científica debería siempre tener en cuenta la potencia. Nadie desea realizar una labor larga y quizá cara de laboratorio o de campo, solo para descubrir tras analizar los datos que los tamaños de la muestra eran insuficientes o que el material experimental era demasiado variable, de forma que los efectos experimentales que se consideraban importantes no se detectaron. Hay dos técnicas disponibles para el investigador al planificar los tamaños adecuados de la muestra. Una técnica es decidir lo pequeño que debería ser cada error típico y escoger n utilizando un análisis como el de la Sección 6.4. Una segunda técnica es un análisis cuantitativo de la potencia del contraste estadístico. Este análisis para el test t se realiza en la Sección 7.7.

Ejercicios 7.3.1-7.3.8

7.3.1 (Ejercicio de muestreo) En referencia a la colección de 100 elipses que se muestra en Ejercicio 3.1.1, que se puede ver como una representación de una población natural del organismo *C. ellipticus*, utilice dígitos aleatorios (de la Tabla 1 o mediante una calculadora) para escoger dos muestras aleatorias de cinco elipses cada una. Utilice una regla para medir la longitud del cuerpo de cada elipse. Basta con tomar medidas con una precisión de milímetro.

- (a) Compare las medias de sus dos muestras, utilizando un test t con $\alpha = 0,05$.
- (b) ¿Produjo el análisis del apartado (a) un error de Tipo I, un error de Tipo II o ningún error?

7.3.2 (Ejercicio de muestreo) Simule que toma muestras aleatorias de dos poblaciones diferentes, como sigue. Primero, proceda como en el Ejercicio 7.3.1 para escoger dos muestras aleatorias de cinco elipses cada una y mida sus longitudes. Después añada 6 mm a *cada* una de las medidas de una de las muestras.

- (a) Compare las medias de sus dos muestras, utilizando un test t con $\alpha = 0,05$.

- (b) ¿Produjo el análisis del apartado (a) un error de Tipo I, un error de Tipo II o ningún error?

7.3.3 (Ejercicio de muestreo) Prepare datos simulados como sigue. Primero, proceda como en el Ejercicio 7.3.1 para tomar dos muestras aleatorias de cinco elipses cada una y mida sus longitudes. Después, lance una moneda. Si sale cara, sume 6 mm a *cada* medida de una de las muestras. Si sale cruz, no modifique ninguna muestra.

- (a) Prepare dos copias de los datos simulados. En la Copia del Estudiante, muestre solo los datos. En la Copia del Instructor, indique también qué muestra (si la hubo) fue modificada.
- (b) Dé su Copia del Instructor al instructor e intercambie su Copia del Estudiante con otro estudiante cuando se le indique.
- (c) Tras haber recibido la copia de otro estudiante, compare las medias de sus dos muestras utilizando un test t de dos colas y $\alpha = 0,05$. Si rechaza H_0 , decida qué muestra fue modificada.

7.3.4 Suponga que se está considerando la aprobación de un nuevo medicamento por la Food and Drug Administra-

tion. La hipótesis nula es que el medicamento no es efectivo. Si la FDA aprueba el medicamento, qué tipo de error, de Tipo I o de Tipo II, posiblemente no se pueda haber cometido?

7.3.5 En el Ejemplo 7.3.1 no se rechazó la hipótesis nula. ¿Qué tipo de error, de Tipo I o de Tipo II podría haberse cometido en ese test t ?

7.3.6 Suponga que se calcula un intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ y su valor es (1,4, 6,7). Si realizamos el contraste $H_0: \mu_1 - \mu_2 = 0$ frente a $H_A: \mu_1 - \mu_2 \neq 0$ utilizando $\alpha = 0,05$, ¿rechazaremos H_0 ? ¿Por qué o por qué no?

7.3.7 Suponga que se calcula un intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ y su valores $(-7,4, -2,3)$. Si realizamos el contraste $H_0: \mu_1 = \mu_2$ frente a $H_A: \mu_1 \neq \mu_2$ utilizando $\alpha = 0,10$, ¿rechazaremos H_0 ? ¿Por qué o por qué no?

7.3.8 Un investigador de productos lácteos ha desarrollado una nueva técnica para fabricar queso que, según afirma, consigue envejecer el queso en sustancialmente menos tiem-

po que los métodos tradicionales, sin afectar a ninguna de sus propiedades. La actualización de las plantas de fabricación de queso con esta nueva tecnología costaría inicialmente millones de dólares, pero si de verdad reduce el tiempo de envejecimiento (incluso marginalmente) producirá altos beneficios para las empresas a largo plazo. Si, por otra parte, el nuevo método no es mejor que el antiguo, la actualización sería un gran error financiero. Antes de tomar la decisión de actualizar, se realizó un experimento para comparar tiempos de fabricación con los métodos nuevo y antiguo.

- Plantee en lenguaje natural cuáles son las hipótesis nula y alternativa de este experimento.
- En el contexto del problema, ¿cuál sería la consecuencia de un error de Tipo I?
- En el contexto del problema, ¿cuál sería la consecuencia de un error de Tipo II?
- En su opinión, ¿qué tipo de error sería más serio? Justifique su respuesta. (Es posible argumentar los dos casos).

7.4 Asociación y causalidad

Cuando comparamos dos poblaciones, a menudo nos enfocamos en la naturaleza de la relación entre una **variable de respuesta** Y (una variable que mide un resultado de interés) y una **variable explicatoria** X (una variable utilizada para explicar o predecir un resultado). Como veremos a continuación, con los datos recogidos de un experimento podemos evaluar si existe o no evidencia de que X afecte al valor medio de Y . Es decir, podemos preguntarnos: ¿causan los cambios en X cambios en Y ? (Por ejemplo, ¿afecta el tolueno a la cantidad media de norepinefrina en el cerebro?). Con los **estudios observacionales**, nuestras conclusiones están más limitadas: no podemos hacer afirmaciones sobre causas, sino más bien solo conclusiones con respecto a la asociación entre X e Y . Por ejemplo, podemos preguntarnos, ¿están los cambios en X asociados con los cambios en el valor medio de Y ? O ¿existe evidencia de que los valores medios de Y son diferentes para las dos poblaciones? (Por ejemplo, ¿tienen los cangrejos de río capturados en dos sitios diferentes longitudes medias distintas?).

Por tanto, la posibilidad de que podamos investigar estas cuestiones depende de cómo se recogieron los datos: experimentalmente o con un estudio observacional. A continuación presentamos ejemplos de cada tipo de estudio en lo que concierne a la comparación de medias de dos muestras, y después realizaremos una presentación más formal de estos tipos de estudio.

Ejemplo 7.4.1 Hematocrito en hombres y mujeres

El nivel de hematocrito es una medida de la concentración de glóbulos rojos en la sangre. La Tabla 7.4.1 presenta las medias y desviaciones típicas muestrales de los valores de hematocrito para dos muestras de jóvenes americanos de 17 años de edad (489 hombres y 469 mujeres)¹⁷. ■

Tabla 7.4.1 Hematocrito (porcentajes)		
	Hombres	Mujeres
Media	45,8	40,6
DT	2,8	2,9

Ejemplo 7.4.2 Pargilina y consumo de sacarosa

Se realizó un estudio para determinar el efecto del medicamento psicoactivo Pargilina sobre el comportamiento en la alimentación del moscardón negro *Phormia regina*. La variable de respuesta fue la cantidad de solución de

sacarosa (azúcar) que una mosca podía beber en 30 minutos. Los investigadores utilizaron dos grupos distintos de moscas: un grupo al que se le había inyectado Pargilina (905 moscas) y un grupo de control al que se le inyectó una solución salina (900 moscas). La comparación de las respuestas de los dos grupos proporciona una evaluación indirecta del efecto de la Pargilina. (Se podría proponer que una forma más *directa* de determinar el efecto de la Pargilina sería medir a cada mosca dos veces: en una ocasión después de inyectarle Pargilina y en otra ocasión después de inyectarle solución salina). Sin embargo, este método directo no es práctico, porque el procedimiento de medida perturba tanto a las moscas que cada mosca solo se puede medir una vez). La Tabla 7.4.2 muestra las medias y las desviaciones típicas de los dos grupos¹⁸. ■

Tabla 7.4.2 Consumo de sacarosa (mg)		
	Control	Pargilina
Media	14,9	46,5
DT	5,4	11,7

En los Ejemplos 7.4.1 y 7.4.2 se realizan comparaciones de dos muestras, pero nótese que los dos estudios se diferencian en un aspecto fundamental. En el Ejemplo 7.4.1 las muestras provienen de una población que aparece naturalmente. El investigador es meramente un observador:

Población 1: valores de hematocrito en hombres americanos de 17 años de edad

Población 2: valores de hematocrito en mujeres americanas de 17 años de edad

Por el contrario, las dos poblaciones del Ejemplo 7.4.2 no existen en la realidad, sino que se definen en términos de condiciones experimentales específicas. En cierto sentido, las poblaciones son creadas por la intervención experimental:

Población 1: consumos de sacarosa de moscardones cuando se les inyecta suero salino

Población 2: consumos de sacarosa de moscardones cuando se les inyecta Pargilina

Estos dos tipos de comparaciones de dos muestras (la observacional y la experimental) se utilizan ampliamente en la investigación. Los métodos formales de análisis son frecuentemente los mismos en los dos tipos, pero la interpretación de los resultados es a menudo algo diferente. Por ejemplo, en el Ejemplo 7.4.2 podría ser razonable decir que la Pargilina *causa* el incremento en el consumo de sacarosa, mientras que esta idea no se podría aplicar en el Ejemplo 7.4.1.

Estudios observacionales frente a estudios experimentales

Una consideración importante al interpretar los resultados de un estudio biológico es si el estudio fue observacional o experimental. En un **experimento**, el investigador interviene o maneja las condiciones experimentales*. En un **estudio observacional**, el investigador simplemente observa una situación existente, como en el ejemplo siguiente.

Ejemplo 7.4.3 Consumo de cigarrillos

En estudios sobre los efectos del consumo de cigarrillos, se ha utilizado tanto el planteamiento experimental como el observacional. Los efectos en animales se pueden estudiar experimentalmente, porque los animales (por ejemplo, los perros) se pueden asignar a grupos de tratamientos y se puede suministrar a cada grupo dosis diferentes de humo de cigarrillo. Los efectos en humanos generalmente se estudian observacionalmente. Por ejemplo, en un estudio, se preguntó a mujeres embarazadas sobre sus hábitos de fumar, hábitos de dieta, etc.¹⁹. Cuando nacieron los bebés, se siguió su desarrollo físico y mental. Un hecho sorprendente relacionado con los pesos al nacer de los bebés: las fumadoras tendían a tener bebés más pequeños que las no fumadoras. La diferencia no fue atribuible al azar (el *P* valor fue menor que 10^{-5}). No obstante, está lejos de estar claro que la diferencia fuera causada por fumar, porque las mujeres que fumaban diferían de las que no fumaban en muchos otros aspectos de su estilo de vida además del hecho de fumar (por ejemplo, tenían hábitos de dieta muy diferentes). ■

* Las condiciones que se manejan deben ser las que definen las poblaciones que se comparan. Por ejemplo, si suministra el mismo medicamento cinco hombres y cinco mujeres y después se comparan los resultados en función del sexo, la comparación entre hombres y mujeres es observacional, no experimental.

Como ilustra el Ejemplo 7.4.3, puede ser difícil determinar la naturaleza exacta de una relación causa-efecto en un estudio observacional. En un experimento, por otra parte, puede ser fácil ver la relación causa-efecto, basándose en la forma en la que el investigador manejó las condiciones experimentales. Para ayudar a fijar las ideas, consideremos un estudio sobre el nivel de colesterol. Supongamos que un grupo de pacientes con niveles altos de colesterol participa en una prueba clínica (es decir, en un experimento médico) en la que algunos de los pacientes escogidos aleatoriamente reciben un nuevo medicamento y otros reciben un medicamento estándar que ha logrado solo efectos modestos en el pasado. Si un test t de dos muestras permite concluir que el nivel medio de colesterol disminuyó más en los que recibieron el nuevo medicamento que en los que recibieron el medicamento estándar, entonces el investigador puede a su vez concluir que el medicamento *causó* un resultado superior y que es mejor que el medicamento estándar.

Consideremos ahora un test t de dos muestras para comparar los niveles medios de colesterol en una muestra aleatoria de personas de 50 años con los niveles medios de colesterol en una muestra aleatoria de personas de 25 años. Supongamos que el test t da un P valor pequeño, a favor de que las personas de 50 años tienen niveles de colesterol mayores que las personas de 25 años. Podemos tener bastante confianza en que el nivel de colesterol tiende a aumentar con la edad. Sin embargo, sería *posible* que existiera alguna otra explicación. Por ejemplo, podría ser que la dieta hubiera cambiado con el tiempo y que las personas de 25 años tomaran alimentos que no toman las personas de 50 años, y que esto fuera la causa de que las personas de 25 años tengan un nivel de colesterol más bajo. Quizá, si las personas de 25 años mantuvieran la misma dieta hasta que tuvieran 50 años, tendrían también niveles bajos de colesterol a la edad de 50.

Como un tercer ejemplo, consideremos la comparación de una muestra aleatoria de propietarios con una muestra aleatoria de inquilinos. Supongamos que se realiza un test t de dos muestras que permite concluir que los propietarios tienen un nivel de colesterol más alto que los inquilinos. No deberíamos concluir que comprar una casa es la causa de que se eleve el nivel de colesterol. Más bien deberíamos considerar que la gente que son dueños de casas tienden a ser mayores que los que las alquilan. Podría darse muy bien el caso de que la edad es el factor causal, que explica por qué los propietarios tienen niveles de colesterol más altos que los inquilinos.

En los tres casos anteriores se realizaría un test t de dos muestras y se rechazaría H_0 . De hecho, podríamos obtener el mismo P valor en los tres. Sin embargo, las conclusiones que podemos extraer de las tres situaciones son bastante diferentes. El alcance de la inferencia que podemos extraer depende de la forma en la que se tomaron los datos. Los experimentos permiten inferir relaciones causa-efecto que en estudios observacionales solo se pueden suponer. Algunas veces, un estudio observacional nos proporcionará una confianza razonable en que entendemos cómo funciona el mecanismo causal. Sin embargo, veremos que la extracción de esas conclusiones está llena de peligros. Por esta razón, los investigadores interesados en extraer conclusiones causales deberían hacer un gran esfuerzo para realizar experimentos controlados en vez de estudios observacionales.

Más sobre estudios observacionales

Las dificultades al interpretar estudios observacionales surgen de dos causas principales:

- Selección no aleatoria de las muestras de las poblaciones.
- Variables externas no controladas.

El ejemplo siguiente ilustra ambos casos.

Ejemplo 7.4.4 Raza y tamaño del cerebro

En el siglo XIX se realizaron muchos esfuerzos en un intento de demostrar «científicamente» que ciertas razas humanas eran inferiores a otras. Un investigador líder en este tema fue el médico americano S. G. Morton, que ganó una amplia admiración por sus estudios sobre el tamaño del cerebro humano. A lo largo de su vida, Morton recogió cráneos humanos de diversas fuentes, y midió cuidadosamente las capacidades craneales de cientos de ellos. Sus datos parecían sugerir que (tal como sospechaba) las razas «inferiores» tenían capacidades craneales inferiores. La Tabla 7.4.3 presenta un resumen de los datos de Morton para la comparación de cráneos caucásicos con cráneos de nativos americanos²⁰. De acuerdo con un test t , la diferencia entre las dos muestras es «estadísticamente significativa» (P valor $< 0,001$). ¿Pero *tiene sentido*?

En primer lugar, la noción de que la capacidad craneal es una medida de inteligencia ya no se toma en serio. Dejando de lado esa cuestión, todavía podemos preguntarnos si es cierto que la capacidad craneal media de los

Tabla 7.4.3 Capacidad craneal (pulgadas ³)		
	Caucásicos	Nativos americanos
Media	87	82
DT	8	10
<i>n</i>	52	144

nativos americanos es menor que la de los caucásicos. Esta inferencia, más allá de los datos reales, requiere que dichos datos se puedan ver como muestras aleatorias de sus poblaciones respectivas. Por supuesto, en la realidad, los datos de Morton no fueron muestras aleatorias sino «muestras por conveniencia», ya que Morton midió aquellos cráneos que pudo obtener. Pero, ¿podrían los datos ser vistos «como si» hubieran sido generados por muestreo aleatorio? Una forma de responder a esta cuestión es buscar las fuentes de sesgo. En 1977, el notorio biólogo Stephen Jay Gould examinó los datos de Morton con este objetivo y, de hecho, Gould encontró varias fuentes de sesgo. Por ejemplo, los 144 cráneos de nativos americanos representaban muchos grupos diferentes de nativos americanos: el 25 % de los cráneos (es decir, 36) provenían de incas peruanos, que son personas de huesos pequeños y cráneos pequeños, mientras que solo relativamente pocos provenían de tribus de cráneos grandes como los Iroquois. Resulta claro que una comparación entre los nativos americanos y los caucásicos no tiene sentido, a menos que de alguna manera se hubieran ajustado esos desequilibrios. Cuando Gould hizo dicho ajuste, obtuvo que la diferencia entre los nativos americanos y los caucásicos desaparecía. ■

Incluso aunque la historia de los cráneos de Morton tiene más de 100 años de antigüedad, puede servir todavía para alertar sobre los riesgos de la inferencia. Morton era un investigador concienzudo y tuvo gran cuidado de realizar medidas precisas. El reexamen de Gould no reveló ningún indicio de fraude consciente en la parte de Morton. Morton pudo haber pasado por alto los sesgos en sus datos porque eran sesgos *invisibles*. Es decir, se relacionaban con aspectos del proceso de selección en vez de con aspectos de las propias medidas.

Cuando observamos un conjunto de datos observacionales, algunas veces podemos quedar tan hipnotizados por su aparente *solidez y objetividad* que podemos olvidar preguntar cómo se seleccionaron las unidades observacionales (las personas o cosas que fueron observadas). Esa pregunta siempre debería realizarse. Si la selección fue caprichosa en vez de verdaderamente aleatoria, los resultados podrían estar severamente distorsionados.

Confusión

Muchos estudios observacionales han tenido como objetivo descubrir algún tipo de relación causal. Esto puede ser algo muy difícil debido a variables externas que intervienen de una forma no controlada (y quizá desconocida). Los investigadores deben tener como guía la siguiente máxima:

Asociación no es causalidad.

Por ejemplo, se sabe que algunas poblaciones cuya dieta es alta en fibra tienen una incidencia reducida de cáncer de colon. Pero esta observación en sí misma no significa que la dieta alta en fibra, y no algún otro factor, sea el que proporciona la protección contra cáncer de colon.

El siguiente ejemplo muestra cómo las variables externas no controladas pueden enturbiar un estudio observacional, y qué tipo de pasos se pueden tomar para clarificar la situación.

Ejemplo 7.4.5 Alas de mariposas

En un estudio observacional grande sobre las mujeres embarazadas, se encontró que las mujeres que fuman cigarrillos tendían a tener bebés más pequeños que las que no fumaban¹⁹. (Este estudio se mencionó en el Ejemplo 7.4.3). Es verosímil que el fumar pueda causar una reducción del peso al nacer, por ejemplo, interfiriendo con el flujo de oxígeno y nutrientes a través de la placenta. Pero, por supuesto, verosimilitud no es demostración. De hecho, los investigadores encontraron que las fumadoras se diferenciaban de las no fumadoras con respecto a muchas otras variables. Por ejemplo, las fumadoras bebían más whisky que las no fumadoras. El consumo de alcohol podría también, de forma verosímil, estar ligado a un déficit en el crecimiento. ■

En el Ejemplo 7.4.5 se presentan tres variables: denominémoslas X = fumar, Y = peso al nacer y Z = consumo de alcohol. Hay una asociación entre X e Y , pero ¿hay una relación causal entre ellas? ¿O hay una relación causal entre Z e Y ? La Figura 7.4.1 presenta una representación esquemática de la situación. Cambios en X están asociados con cambios en Y . Sin embargo, cambios en Z están también asociados con cambios en Y . Se dice que el efecto que X tiene sobre Y está **confundido** con el efecto que Z tiene sobre Y . En el contexto del Ejemplo 7.4.5 se dice que el efecto de fumar sobre el peso al nacer está confundido con el efecto del consumo de alcohol sobre el peso al nacer. En estudios observacionales, la confusión de efectos es un problema común.

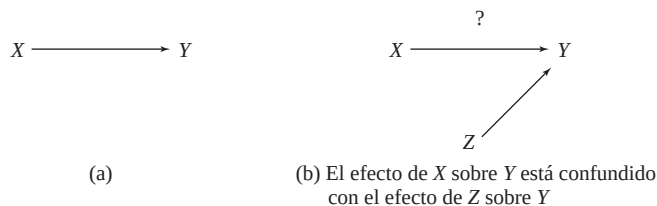


Figura 7.4.1 Representación esquemática de la causalidad (a) y de la confusión (b)

Ejemplo 7.4.6 Fumar y peso al nacer

El estudio presentado en el Ejemplo 7.4.5 permitió descubrir muchas variables de confusión. Por ejemplo, las fumadoras bebían más café que las no fumadoras. Además (y esto es especialmente intrigante) se encontró que las fumadoras tenían su primera menstruación a edades más tempranas que las no fumadoras. Este fenómeno (presentación más temprana de la primera menstruación) posiblemente no podría haber sido causado por el fumar, porque sucedía (en casi todos los casos) *antes* de que la mujer empezara a fumar. Una interpretación que ha sido propuesta es que las dos poblaciones (mujeres que escogieron fumar y mujeres que no escogieron fumar) son diferentes en algún aspecto biológico. Por tanto, se ha sugerido que el peso reducido al nacer es debido «a la *fumadora*, no al *fumar*»²¹.

Diversos estudios más recientes han intentado arrojar alguna luz sobre la relación entre el hábito de fumar materno y el desarrollo del bebé. Unos investigadores en un estudio observaron, además del hábito de fumar, aproximadamente 50 variables externas, entre las que se encontraban la edad de la madre, el peso, la altura, el grupo sanguíneo, la circunferencia superior del brazo, la religión, la educación, los ingresos, etc.²². Tras aplicar métodos estadísticos complejos de ajuste, concluyeron que el peso al nacer varía con el hábito de fumar incluso cuando estos factores externos se mantienen constantes. Esto indica que existe bastante verosimilitud de una relación entre X = fumar e Y = peso al nacer, como se muestra en la Figura 7.4.1, aunque algunas otras variables pudieran afectar también al peso al nacer. El asunto es que la presencia de confusión no significa que no exista una relación entre X e Y , sino solo que está mezclada con otros efectos, por lo que tenemos que ser precavidos al interpretar los resultados de un estudio observacional.

En otro estudio sobre mujeres embarazadas, unos investigadores midieron varias magnitudes relacionadas con el funcionamiento de la placenta²³. Obtuvieron que, comparadas con las no fumadoras, las mujeres que fumaban tenían más anomalías en la placenta, y que sus niños tenían niveles mucho más altos de cotinina, una sustancia derivada de la nicotina. Encontraron también evidencia de que, en las mujeres que fumaban, la circulación de la sangre en la placenta aumentaba notablemente tras abstenerse de fumar durante tres horas.

Un tercer estudio utilizó un diseño ajustado para intentar aislar el efecto del hábito de fumar. Los investigadores identificaron 159 mujeres que habían fumado durante un embarazo pero que habían dejado de fumar antes del siguiente embarazo²⁴. Estas mujeres se emparejaron individualmente con 159 mujeres que fumaron durante dos embarazos consecutivos. Las parejas se ajustaron con respecto al peso al nacer del primer hijo, la cantidad que habían fumado durante el primer embarazo y varios otros factores. Por tanto, los miembros de una pareja se suponían con idéntico «potencial reproductivo». Los investigadores consideraron después el peso al nacer del segundo hijo. Obtuvieron que las mujeres que habían dejado de fumar dieron a luz bebés que pesaban más que los de su respectiva pareja que continuó fumando. Por supuesto, no podemos excluir la posibilidad de que las mujeres que dejaron de fumar dejaran también otros hábitos dañinos, como el beber demasiado alcohol, y que el aumento en el peso al nacer no fuera realmente debido a dejar de fumar. ■

El Ejemplo 7.4.6 demuestra que los estudios observacionales pueden proporcionar información sobre causalidad, pero que debe ser interpretada con cuidado. Los investigadores generalmente están de acuerdo en que una interpretación causal de una asociación observada requiere un soporte adicional (por ejemplo, que la asociación sea observada de forma consistente en estudios observacionales realizados bajo diversas condiciones y teniendo en cuenta diversos factores externos), y también, idealmente, que la relación causal sea sostenida por la evidencia experimental. No queremos decir que una asociación observada no pueda ser interpretada causalmente, sino solo que esa interpretación requiere un especial cuidado.

Asociación engañosa

Ejemplo 7.4.7 Ultrasonidos

Es bastante común que los médicos examinen por ultrasonidos el feto de una mujer embarazada. Sin embargo, cuando se utilizó por primera vez la tecnología de ultrasonidos, había preocupación por si el procedimiento podría causar daño al bebé. Un estudio inicial pareció corroborarlo. En promedio, los bebés expuestos a los ultrasonidos en el útero pesaban menos al nacer que los bebés no expuestos a ultrasonidos²⁵. Posteriormente, se realizó un estudio en el que algunas mujeres elegidas aleatoriamente recibieron ultrasonidos y otras no. Este estudio no encontró diferencia en los pesos al nacer entre los dos grupos²⁶. Parece que la razón de la diferencia que apareció en el primer estudio era que los ultrasonidos se utilizaban más frecuentemente en mujeres que experimentaban problemas en el embarazo. Las complicaciones en el embarazo, no el uso de ultrasonidos, eran la causa del menor peso al nacer. ■

La Figura 7.4.2 es una representación esquemática de la situación del Ejemplo 7.4.7. Los cambios en X (haber sido examinada por ultrasonidos) se asocian con cambios en Y (menor peso al nacer). Sin embargo, X e Y dependen ambas de una tercera variable Z (si ha habido o no ha habido problemas en el embarazo), que es la variable que controla la relación. Los cambios en X y en Y son una respuesta común a la tercera variable Z . Se dice en este caso que la asociación entre X e Y es **engañosa**. Cuando controlamos la «variable oculta» Z , la relación entre X e Y desaparece. En el caso del Ejemplo 7.4.7, no era el haber recibido ultrasonidos lo que influía en el peso al nacer. Lo que importaba era si había habido o no había habido problemas en el embarazo.

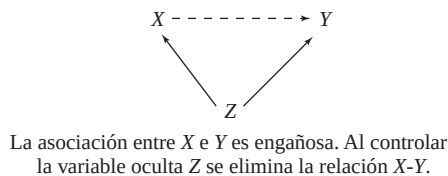


Figura 7.4.2 Representación esquemática de una asociación engañosa

Más sobre experimentos

Un experimento es un estudio en el que el investigador interviene e impone las condiciones del tratamiento. A continuación se presenta un ejemplo simple.

Ejemplo 7.4.8 Dolor de cabeza

Suponga que un investigador da ibuprofeno a algunas personas que tienen dolores de cabeza y aspirina a otras, y después mide el tiempo que tarda en desaparecer el dolor de cabeza en cada persona. En este caso, hay dos tratamientos: ibuprofeno y aspirina. Al asignar las personas a los grupos de tratamiento (ibuprofeno y aspirina) el investigador está dirigiendo el experimento. ■

Cuando estamos planteando un experimento, las unidades a las que se asignan los tratamientos se denominan **unidades experimentales**. En un experimento sobre agricultura, la unidad experimental podría ser una parcela de tierra. En general, la unidad experimental es la mínima unidad a la que se aplica un tratamiento en un experimento. Por tanto, en el Ejemplo 7.4.8, las unidades experimentales son las personas, ya que el tratamiento se asigna persona a persona.

Si los tratamientos se asignan aleatoriamente, por ejemplo, lanzando una moneda y si sale cara la persona recibe ibuprofeno, mientras que si sale cruz la persona recibe aspirina, entonces el experimento es un experimento *aleatorizado*. Supongamos que se realiza un experimento en el que un grupo recibe un tratamiento y otro grupo (el grupo de control) no recibe nada. Por ejemplo, se podría investigar la efectividad del ibuprofeno en el tratamiento del dolor de cabeza dándolo a algunas personas, y a otras no dándoles nada. Por el contrario, se dice que el experimento en el que algunas personas reciben ibuprofeno y otras reciben aspirina tiene un control «activo» (el grupo de la aspirina).

Distribuciones en la aleatorización

En la Sección 5.2 desarrollamos el concepto de distribución muestral de la media muestral, $\bar{Y}$, considerando como variable $\bar{Y}$ de una muestra aleatoria a otra. Estrictamente hablando, esto proporciona la base de la inferencia cuando se analiza un estudio observacional, pero no cuando los datos surgen de un experimento (en el que los tratamientos se asignan a unidades experimentales, en vez de a una muestra aleatoria que se toma de una población). Sin embargo, los conceptos de la Sección 5.2 se pueden ampliar de forma natural para definir la **distribución en la aleatorización** de $\bar{Y}$, que es la distribución que toma $\bar{Y}$ en todas las posibles asignaciones aleatorias dentro de un experimento. Las distribuciones en la aleatorización son entonces la base para realizar inferencia a partir de experimentos.

¿Solo estadística?

El término «estadística» se utiliza algunas veces (o, más bien, se confunde) con un epíteto. Por ejemplo, algunas personas dicen que la evidencia que relaciona el colesterol en la dieta con la enfermedad cardíaca es «solamente estadística». Lo que realmente quieren decir es que es «solamente observacional». La evidencia estadística puede ser muy fuerte en todo caso, si surge de un experimento aleatorizado en vez de provenir de un estudio observacional. Como molestan los ejemplos anteriores, la evidencia estadística de un estudio observacional se debe interpretar con gran cuidado, debido a las distorsiones potenciales causadas por variables externas.

Ejercicios 7.4.1-7.4.9

7.4.1 En 2005, el 5,3 % de las muertes de Estados Unidos fueron causadas por enfermedades crónicas del tracto respiratorio bajo (por ejemplo, asma y enfisema). En Arizona, el 6,2 % de las muertes fueron debidas a enfermedades crónicas del tracto respiratorio bajo²⁷. ¿Significa esto que vivir en Arizona exacerba los problemas respiratorios? Si no es así, ¿cómo se puede explicar que el porcentaje de Arizona esté por encima del porcentaje nacional?

7.4.2 Existe la hipótesis de que los implantes de silicona en el pecho causan enfermedades. En un estudio se encontró que era más probable que las mujeres con implantes fumaran, bebieran mucho, utilizaran tinte para el pelo y hubieran tenido un aborto, con respecto a las mujeres del grupo de comparación que no tenían implantes²⁸. Utilice el lenguaje de la estadística para explicar por qué este estudio produce dudas sobre la afirmación de que los implantes causan enfermedades.

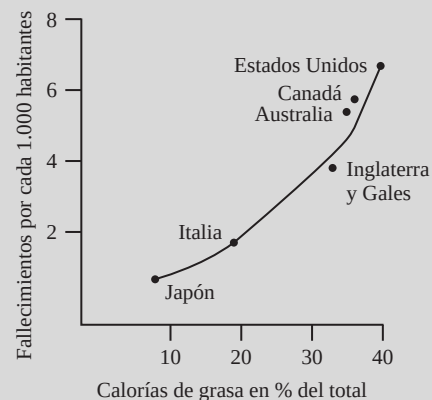
7.4.3 Considere el planteamiento del Ejercicio 7.4.2.

- (a) ¿Cuál es la variable explicativa?
- (b) ¿Cuál es la variable de respuesta?
- (c) ¿Cuáles son las unidades observacionales?

7.4.4 En un estudio sobre 1.040 sujetos, unos investigadores encontraron que la preponderancia de enfermedades coronarias del corazón aumentaba con el incremento del número de tazas de café consumidas por día²⁹.

- (a) ¿Cuál es la variable explicativa?
- (b) ¿Cuál es la variable de respuesta?
- (c) ¿Cuáles son las unidades observacionales?

7.4.5 En un estudio preliminar sobre la relación entre la dieta y las enfermedades del corazón, los investigadores obtuvieron datos sobre mortalidad por enfermedades cardíacas en varios países y también sobre las composiciones promedio de las dietas en los mismos países. La gráfica que se presenta a continuación muestra, para seis países, la tasa de muertes por enfermedades cardíacas degenerativas entre



1948 y 1949 (en hombres entre 55 y 59 años de edad), en función de la cantidad de grasa en la dieta³⁰.

¿De qué formas podría ser engañosa esta gráfica? ¿Qué variables externas podrían ser relevantes en este caso? Comente sus respuestas.

7.4.6 Un poco antes del día de los enamorados en 1999, un artículo de un periódico tenía el titular siguiente: «Estudios demuestran que el matrimonio proporciona una vida más larga y más saludable». El titular se basó en estudios que habían demostrado que las personas casadas vivían más y tenían menos tasas de cáncer, enfermedades cardíacas y derrames cerebrales que aquellos que nunca se habían casado³¹. Utilice el lenguaje de la estadística para discutir el titular. Utilice un diagrama esquemático similar al de la Figura 7.4.1 o al de la Figura 7.4.2 para fundamentar su explicación de la situación.

7.4.7 En junio de 2009, el *New York Times* publicó un artículo titulado «¿El alcohol es bueno para usted? Algunos científicos lo dudan». El autor escribía, «Estudio tras estudio sugieren que tomar alcohol con moderación puede fomentar la salud del corazón e incluso prevenir contra la diabetes y la demencia. La evidencia es tan abundante que algunos expertos consideran que beber moderadamente (aproximadamente una copa al día para las mujeres y aproximadamente dos para los hombres) puede ser un componente central de un estilo de vida saludable». Posteriormente en el artículo, el autor escribía: «Para algunos científicos, la cuestión no admite discusión. Ningún estudio, dicen estos críticos, ha probado nunca una relación causal entre beber moderadamente y un menor riesgo de muerte». Explique, utilizando el lenguaje de la estadística y un diagrama esquemático similar al de la Figura 7.4.1 o al de la Figura 7.4.2, por qué los críticos dicen que ningún estudio ha demostrado nunca una relación causal.

7.4.8 En un estudio sobre la relación entre el peso al nacer y la raza, se examinaron registros con nacimientos de bebés en Illinois. Los investigadores encontraron que el porcentaje de bebés con bajo peso al nacer entre bebés nacidos en Estados Unidos de mujeres blancas era mucho menor que el porcentaje de bebés con bajo peso al nacer entre los bebés nacidos en Estados Unidos de mujeres negras. Esto sugiere que la raza tiene un papel importante en determinar la probabilidad de que un bebé tenga un bajo peso al nacer. Sin embargo, el porcentaje de bebés con bajo peso al nacer entre los bebés nacidos en África de mujeres negras era aproximadamente igual al porcentaje de los bebés nacidos en Estados Unidos de mujeres blancas³². Utilice el lenguaje de la estadística para comentar lo que dicen estos datos sobre las relaciones entre el bajo peso al nacer, la raza, y el lugar donde la

madre da a luz. Utilice un diagrama esquemático similar al de la Figura 7.4.1 o al de la Figura 7.4.2 para fundamentar su explicación.

7.4.9 ¿Causa el lanzamiento de un libro de Harry Potter que los niños dediquen más tiempo a leer y, por lo tanto, se reduzca el número de accidentes que tienen? Unos doctores en Inglaterra compararon el número de visitas a las urgencias de un hospital debidas a daños musculares o de huesos en niños entre 7 y 15 años de edad, durante dos tipos de fines de semana: (1) después de la fecha de lanzamiento de dos libros de Harry Potter y (2) durante 24 fines de semana de «control». La tabla que se presenta a continuación muestra los datos, con los «fines de semana de Harry Potter» en cursiva³³.

Fin de semana	Daños	Fin de semana	Daños
6/7/03	63	7/10/04	57
6/14/03	77	7/17/04	66
6/21/03	36*	7/24/04	62
6/28/03	63	6/4/05	51
7/5/03	75	6/11/05	83
7/12/03	71	6/18/05	60
7/19/03	60	6/25/05	66
7/26/03	52	7/2/05	74
6/5/04	78	7/9/05	75
6/12/04	84	7/16/05	37*
6/19/04	70	7/23/05	46
6/26/04	75	7/30/05	68
7/3/04	81	8/6/05	60

- Dada la naturaleza de los datos, ¿podemos realizar una inferencia sobre que el lanzamiento de los libros de Harry Potter causa un cambio en los accidentes? ¿Por qué o por qué no?
- El promedio de los fines de semana de Harry Potter es 36,5, con una desviación típica de 0,7. Los valores correspondientes para los otros fines de semana (de control) son 67,4 y 10,4. Utilice un test t para investigar la afirmación de que el número pequeño de daños durante los fines de semana de Harry Potter es consistente con una variación por azar. Utilice $\alpha = 0,01$. [Nota: la fórmula (6.7.1) da un valor de 23,9° de libertad para estos datos].

7.5 Tests t de una cola

El test t descrito en las secciones anteriores se denomina **test t de dos colas** o **test t bilateral**, porque la hipótesis nula se rechaza si t_s está en una de las colas de la distribución t de Student y el P valor de los datos es un área de dos colas bajo la curva t de Student. Un test de dos colas se utiliza para comprobar la hipótesis nula

$$H_0: \mu_1 = \mu_2$$

frente a la hipótesis alternativa

$$H_A: \mu_1 \neq \mu_2$$

Esta alternativa H_A se denomina **alternativa no direccional**.

Hipótesis alternativas direccionales

En algunos estudios es aparente desde el principio (*antes* de tomar los datos) que solo hay una dirección razonable de desviación con respecto a H_0 . En estas situaciones resulta apropiado formular una hipótesis alternativa direccional. La siguiente es una alternativa direccional:

$$H_A: \mu_1 < \mu_2$$

Otra alternativa direccional es

$$H_A: \mu_1 > \mu_2$$

Los dos ejemplos siguientes ilustran situaciones dónde las alternativas direccionales resultan apropiadas.

Ejemplo 7.5.1 Suplemento de niacina

Consideremos un experimento de alimentación con corderos. La observación Y será la ganancia de peso durante un experimento de dos semanas. 10 animales recibirán la dieta 1 y 10 animales recibirán la dieta 2, siendo

Dieta 1 = Ración estándar + niacina

Dieta 2 = Ración estándar

Se puede esperar, sobre bases biológicas, que la niacina aumente la ganancia de peso. No hay ninguna razón para pensar que pueda disminuir la ganancia de peso. Una formulación apropiada sería

H_0 : La niacina no es efectiva para aumentar la ganancia de peso ($\mu_1 = \mu_2$).

H_A : La niacina es efectiva para aumentar la ganancia de peso ($\mu_1 > \mu_2$). ■

Ejemplo 7.5.2 Tinte para el pelo y cáncer

Supongamos que se prueba un cierto tinte para el pelo para determinar si es carcinógeno (causa cáncer). El tinte se aplica en la piel de 20 ratones (grupo 1), y se aplica una sustancia inerte en la piel de otros 20 ratones (grupo 2) que sirven como control. La observación Y será el número de tumores que aparecen en cada ratón. Una formulación apropiada es

H_0 : El tinte no es carcinógeno ($\mu_1 = \mu_2$).

H_A : El tinte es carcinógeno ($\mu_1 > \mu_2$). ■

Nota: Si H_A es direccional, entonces algunas personas reformulan H_0 para incluir la dirección opuesta. Por ejemplo, si H_A es $H_A: \mu_1 > \mu_2$, entonces podríamos expresar H_0 como $H_0: \mu_1 \leq \mu_2$. Por tanto, la hipótesis nula indica que la media de la población 1 *no* es mayor que la media de la población 2, mientras que la hipótesis alternativa asevera que la media de la población 1 es mayor que la media de la población 2. Con estas dos hipótesis, se cubren todas las posibilidades.

Procedimiento del test de una cola

Cuando la hipótesis alternativa es direccional, el procedimiento del test t se debe modificar. El procedimiento modificado se denomina **test t de una cola** y se realiza en dos pasos como sigue:

Paso 1 Comprobar la direccionalidad. Ver si los datos se desvían con respecto a H_0 en la dirección especificada por H_A :

- (a) Si no es así, el P valor es mayor que 0,50.
- (b) Si es así, seguir con el paso 2.

Paso 2 El P valor es el área en una cola más allá de t_s .

Para concluir el test, se realiza una decisión con un nivel de significación preestablecido α : H_0 se rechaza si el P valor $\leq \alpha$.

La base para el procedimiento en dos pasos es que el P valor mide la desviación con respecto a H_0 en la dirección especificada por H_A . El P valor de una cola se ilustra en la Figura 7.5.1 para los dos casos en los que los datos se desvían con respecto a H_0 en la dirección especificada por H_A . La Figura 7.5.2 ilustra el P valor para (a) un caso en el que los datos son consistentes con $H_A: \mu_1 > \mu_2$, y (b) un caso en el que los datos son inconsistentes con $H_A: \mu_1 > \mu_2$. El procedimiento de contraste de dos pasos se ilustra en el Ejemplo 7.5.3.

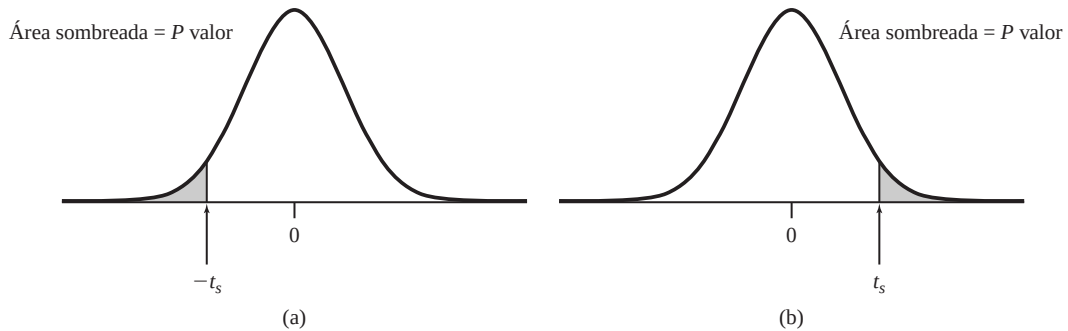


Figura 7.5.1 P valor de una cola para un test t , (a) si la alternativa es $H_A: \mu_1 < \mu_2$ y t_s es negativo; (b) si la alternativa es $H_A: \mu_1 > \mu_2$ y t_s es positivo

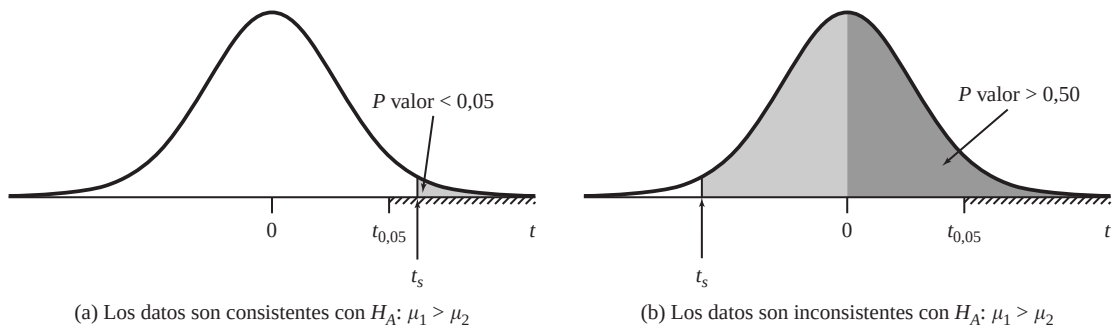


Figura 7.5.2 P valor de una cola para un test t , (a) en el que los datos son consistentes con $H_A: \mu_1 > \mu_2$; (b) en el que los datos son inconsistentes con $H_A: \mu_1 > \mu_2$

Ejemplo 7.5.3 Suplemento de niacina

Considere el experimento de alimentación de corderos del Ejemplo 7.5.1. La hipótesis alternativa es

$$H_A: \mu_1 > \mu_2$$

Afirmaremos que existe evidencia significativa a favor de H_A si $\bar{Y}_1$ es suficientemente mayor que $\bar{Y}_2$. Supongamos que la fórmula (6.7.1) da como resultado $gl = 18$. La Tabla 7.5.1 reproduce los valores críticos de la Tabla 4.

Tabla 7.5.1 Valores críticos con $gl = 18$										
Área de la cola	0,20	0,10	0,05	0,04	0,03	0,025	0,02	0,01	0,005	0,0005
Valor crítico	0,862	1,330	1,734	1,855	2,007	2,101	2,214	2,552	2,878	3,922

Para ilustrar el procedimiento del test de una cola, supongamos que tenemos³⁴

$$ET_{(\bar{Y}_1 - \bar{Y}_2)} = 2,2 \text{ lb}$$

y que elegimos $\alpha = 0,05$. Consideremos varias posibilidades para las dos medias muestrales.

- (a) Supongamos que los datos son $\bar{y}_1 = 10$ lb e $\bar{y}_2 = 13$ lb. Esta desviación con respecto a H_0 es opuesta a la aseveración de H_A . Tenemos que $\bar{y}_1 < \bar{y}_2$, pero H_A asevera que $\mu_1 > \mu_2$. En consecuencia, P valor $> 0,50$, por lo que no encontraríamos evidencia significativa a favor de H_A con ningún nivel de significación. (Nunca se utiliza un valor de α mayor que 0,50). Concluimos que los datos no proporcionan evidencia de que la niacina sea efectiva para aumentar la ganancia de peso.
- (b) Supongamos que los datos son $\bar{y}_1 = 14$ lb e $\bar{y}_2 = 10$ lb. Esta desviación con respecto a H_0 es en la dirección de H_A (porque $\bar{y}_1 > \bar{y}_2$), por lo que procedemos con el paso 2. El valor de t_s es

$$t_s = \frac{(14 - 10) - 0}{2,2} = 1,82$$

El P valor (de una cola) del test es la probabilidad de obtener un estadístico t , con 18 grados de libertad, que sea tan grande como o mayor que 1,82. Esta probabilidad de la cola superior (obtenida con un computador) vale 0,043, como se muestra en la Figura 7.5.3.

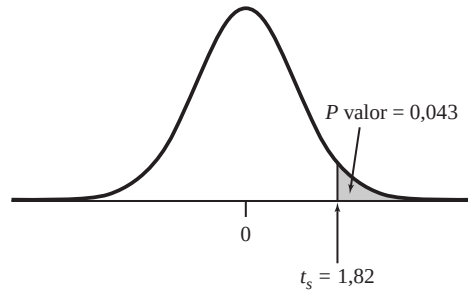


Figura 7.5.3 P valor de una cola para el test t del Ejemplo 7.5.3

Si no tenemos un computador o una calculadora gráfica disponibles, podríamos utilizar la Tabla 4 para acotar el P valor. Utilizando dicha Tabla 4, podemos ver que el P valor estaría acotado como sigue:

$$0,04 < P \text{ valor de una cola} < 0,05$$

Como el P valor $< \alpha$, rechazamos H_0 y concluimos que existe alguna evidencia de que la niacina es efectiva.

- (c) Supongamos que los datos son $\bar{y}_1 = 11$ lb e $\bar{y}_2 = 10$ lb. Entonces, procediendo como en el apartado (b), calculamos el estadístico de contraste y su valor es $t_s = 0,45$. El P valor es 0,329.

Si no tenemos un computador o una calculadora gráfica disponibles, podríamos utilizar la Tabla 4 para acotar el P valor como

$$P \text{ valor} > 0,20$$

Como el P valor $> \alpha$, no encontramos evidencia significativa a favor de H_A . Concluimos por tanto que no hay evidencia suficiente para afirmar que la niacina es efectiva. Por tanto, aunque estos datos se desvían con respecto a H_0 en la dirección de H_A , la cantidad de desviación no es lo suficientemente grande como para justificar una evidencia significativa a favor de H_A . ■

Nótese que lo que distingue un test t de una cola de un test t de dos colas es la forma en la que se determina el P valor, pero no en la direccionalidad o la falta de direccionalidad de la conclusión. Si obtenemos evidencia significativa a favor de H_A , entonces la conclusión se puede considerar direccional incluso aunque nuestra H_A no sea direccional. (Por ejemplo, en el Ejemplo 7.2.4 concluimos que el tolueno aumenta la concentración de NE).

Alternativa direccional frente a alternativa no direccional

Los mismos datos proporcionarán un P valor diferente dependiendo de si la hipótesis alternativa es direccional o no direccional. De hecho, si los datos se desvían respecto a H_0 en la dirección especificada por H_A , el P valor de una hipóte-

sis alternativa direccional será 1/2 del P valor de un test que utilice una hipótesis no direccional. Puede suceder que los mismos datos proporcionen evidencia significativa a favor de H_A utilizando el procedimiento de una cola pero no utilizando el procedimiento de dos colas, como muestra el Ejemplo 7.5.4.

Ejemplo 7.5.4 Suplemento de niacina

Considere el apartado (b) del Ejemplo 7.5.3. En dicho ejemplo escogimos $\alpha = 0,05$ y contrastamos

$$H_0: \mu_1 = \mu_2$$

frente a la hipótesis alternativa direccional

$$H_A: \mu_1 > \mu_2$$

Con $\bar{y}_1 = 14$ lb e $\bar{y}_2 = 10$ lb, el estadístico de contraste varía $t_s = 1,82$ y el P valor era 0,043, tal como se indicaba en la Figura 7.5.3. Nuestra conclusión fue afirmar que había evidencia significativa a favor de H_A .

Sin embargo, supongamos que hubiéramos deseado contrastar

$$H_0: \mu_1 = \mu_2$$

frente a la hipótesis alternativa no direccional

$$H_A: \mu_1 \neq \mu_2$$

Con los mismos datos de $\bar{y}_1 = 14$ lb e $\bar{y}_2 = 10$ lb, el estadístico de contraste tiene el mismo valor de $t_s = 1,82$. El P valor, sin embargo, es 0,086, como se muestra en la Figura 7.5.4. Por tanto, P valor $> \alpha$ y no rechazamos H_0 .

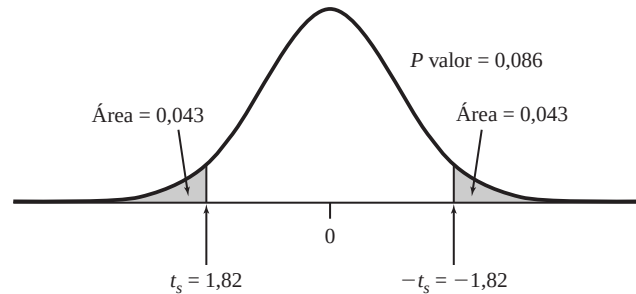


Figura 7.5.4 P valor de dos colas para el test t del Ejemplo 7.5.4

Por tanto, el procedimiento de una cola encuentra evidencia significativa a favor de H_A , pero el procedimiento de dos colas no. En este sentido, es «más fácil» afirmar que la evidencia significativa sostiene a la hipótesis H_A con el procedimiento de una cola que con el procedimiento de dos colas. ■

¿Por qué el P valor para dos colas se reduce a la mitad cuando la hipótesis alternativa es direccional? En el Ejemplo 7.5.4, el investigador concluiría diciendo «los datos sugieren que la niacina incrementa la ganancia de peso. Pero si la niacina no tiene efecto, entonces el tipo de datos que obtuve en mi experimento, es decir, tener dos medias muestrales que se diferencian en 1,82 ET o más, podría aparecer bastante a menudo (P valor = 0,086). Algunas veces, la dieta de niacina parecería mejor; otras veces la dieta estándar parecería mejor. No puedo encontrar evidencia significativa a favor en la base a lo que visto en estos datos». En el Ejemplo 7.5.3(b), el investigador concluiría diciendo «antes de realizar el experimento, yo sospechaba que la niacina aumenta la ganancia de peso. Los datos proporcionan evidencia que sostiene esta teoría. Si la niacina no tiene efecto, entonces el tipo de datos que he obtenido en mi experimento, es decir, que la media muestral de la dieta de niacina supera a la dieta estándar en 1,82 ET o más, sucedería raramente (P valor = 0,043). (Antes de realizar el experimento yo desestimé la posibilidad de que la media de la dieta de niacina pudiera ser menor que la media de la dieta estándar). Por tanto, puedo afirmar que tengo evidencia significativa que sostiene a la hipótesis H_A ». El investigador del Ejemplo 7.5.3(b) está utilizando dos fuentes de información para afirmar la significación de la evidencia a favor de H_A : (1) lo que los datos tienen que decir (medido por el área de la cola) y (2) expectativas previas

(que permiten al investigador ignorar el área de la cola inferior, es decir, el área de 0,043 por debajo de la curva hacia la izquierda de $-1,82$ en la Figura 7.5.4).

Nótese que la modificación del procedimiento, cuando se pasa de un test de dos colas a un test de una cola, mantiene la interpretación del nivel de significación α que se dio en la Sección 7.3, es decir,

$$\alpha = \Pr\{\text{rechazar } H_0\} \text{ si } H_0 \text{ es cierta}$$

Por ejemplo, consideremos el caso de $\alpha = 0,05$. La Figura 7.5.5 muestra que el área total sombreada (la probabilidad de rechazar H_0) es igual a 0,05 tanto en un test de dos colas como en un test de una cola. Esto significa que, si un gran número de investigadores contrastaran una hipótesis H_0 verdadera, entonces el 5 % de ellos obtendrían evidencia significativa a favor de H_A y cometerían un error de Tipo I. Esta afirmación es verdadera tanto si la alternativa es direccional como si es no direccional.

El punto crucial en la justificación del procedimiento modificado para contrastar contra una hipótesis H_A direccional es que si la dirección de desviación de los datos con respecto a H_0 no es la especificada por H_A , entonces no afirmaremos que la evidencia sostiene de forma significativa a la hipótesis H_A . Por ejemplo, en el experimento sobre carcinogénesis del Ejemplo 7.5.2, si los ratones expuestos al tinte del pelo tuvieran *menos* tumores que el grupo de control podríamos (1) concluir simplemente que los datos no indican un efecto carcinógeno, o (2) si el grupo expuesto tuviera *sustancialmente* menos tumores, de forma que el estadístico estuviera muy lejos en la cola incorrecta de la distribución t , podríamos buscar errores metodológicos del experimento (por ejemplo, problemas en las técnicas de laboratorio o en la grabación de los datos, en la asignación aleatoria de los ratones a los dos grupos, etc.) pero no afirmaríamos que existe evidencia significativa a favor de H_A .

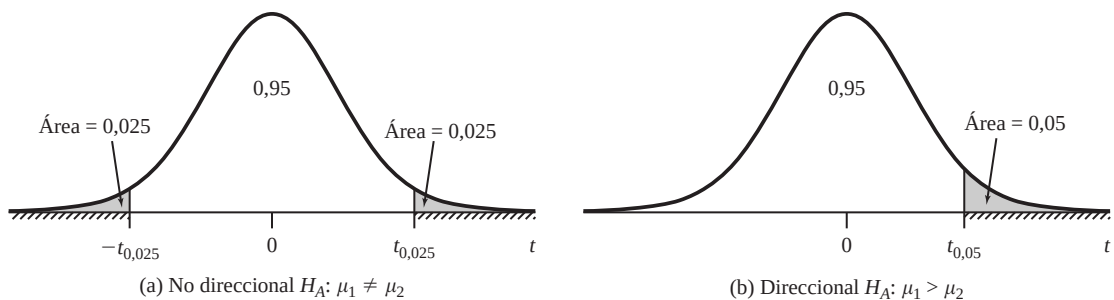


Figura 7.5.5 Test t de dos colas y de una cola con $\alpha = 0,05$. Los datos proporcionan evidencia significativa a favor de si H_A está en la región rayada del eje t

El uso de un test t de una cola resulta especialmente natural cuando solo es plausible una dirección de desviación con respecto a H_0 . Sin embargo, los tests de una cola se utilizan también en situaciones donde es posible la desviación en ambas direcciones, pero solo una dirección es de interés. Por ejemplo, en el experimento sobre la niacina del Ejemplo 7.5.3, no es necesario que el investigador piense que es *imposible* que la niacina reduzca la ganancia de peso en vez de aumentarla. Las desviaciones en la dirección incorrecta (menos ganancia de peso con niacina) no conducirían a afirmar que hay evidencia significativa a favor de H_A y, por tanto, no haríamos afirmaciones sobre el efecto de la niacina. Este es el aspecto esencial que distingue una formulación direccional de una no direccional.

Elección de la forma de H_A

¿Cuándo es legítimo utilizar una H_A direccional y, por tanto, realizar un test de una cola? La respuesta a esta pregunta está relacionada con la comprobación sobre la direccionalidad (paso 1 del procedimiento de contraste en dos pasos que se presentó previamente). Claramente, esta comprobación tiene sentido solo si H_A se formula antes de inspeccionar los datos. (Si formuláramos una H_A direccional que estuviera «inspirada» por los datos, entonces, por supuesto, los datos siempre se desviarían con respecto a H_0 en la dirección «correcta» y el procedimiento de contraste siempre procedería al paso 2). Esta es la base para la siguiente regla.

Regla para alternativas direccionales

Es legítimo utilizar una alternativa H_A direccional solo si H_A se formula antes de ver los datos y no hay interés científico en los resultados que se desvíen en la dirección opuesta a la especificada por H_A .

Al realizar su trabajo, los investigadores se encuentran a menudo más satisfechos cuando encuentran evidencia significativa a favor de una hipótesis alternativa que cuando no la encuentran. De hecho, los informes de investigación contienen a menudo frases como «no pudimos obtener evidencia significativa a favor de la hipótesis alternativa» o «los resultados no proporcionaron significación estadística». Bajo estas circunstancias, podríamos preguntarnos cuáles serían las consecuencias si los investigadores sucumbieran a la tentación natural de ignorar la regla anterior sobre la utilización de alternativas direccionales. Después de todo, es muy frecuente que se pueda pensar en el fundamento de un efecto *ex post facto* (es decir, después de que un efecto ha sido observado). Una vuelta el experimento imaginario sobre los gustos musicales de las plantas ilustrará esta situación.

Ejemplo 7.5.5 Música y caléndulas

Recordemos el experimento imaginario del Ejemplo 7.3.2, en el que los investigadores medían las alturas de caléndulas expuestas a la música de Bach o de Mozart. Supongamos, como antes, que la hipótesis nula es cierta, que $g_l = 60$ y que todos los investigadores realizaron tests t con $\alpha = 0,05$. Supongamos ahora adicionalmente que todos los investigadores violan la regla de utilizar alternativas direccionales, y que formulan H_A después de ver los datos. La mitad de los investigadores obtendrían datos para los que $\bar{y}_1 > \bar{y}_2$, y formularían la alternativa

$$H_A: \mu_1 > \mu_2 \text{ (las plantas prefieren a Bach)}$$

La otra mitad obtendría datos para los que $\bar{y}_1 < \bar{y}_2$, y formularían la alternativa

$$H_A: \mu_1 < \mu_2 \text{ (las plantas prefieren a Mozart)}$$

Observemos ahora lo que sucedería. Como los investigadores están utilizando alternativas direccionales, todos calcularían P valores utilizando solamente una cola de la distribución. Podríamos esperar que ocurriera lo siguiente:

- El 90 % de ellos obtendría un valor de t_s en el 90 % medio de la distribución y no obtendrían evidencia significativa a favor de H_A .
- El 5 % de ellos obtendría un valor de t_s en el 5 % superior de la distribución y concluirían que las plantas prefieren a Bach.
- El 5 % de ellos obtendría un valor de t_s en el 5 % inferior de la distribución y concluirían que las plantas prefieren a Mozart.

Por tanto, un total del 10 % de los investigadores afirmaría que existe evidencia significativa a favor de H_A . Por supuesto, cada investigador por separado nunca se daría cuenta de que el porcentaje total de errores de Tipo I es del 10 %, en vez del 5 %. Y las conclusiones de que las plantas prefieren a Bach o Mozart se podrían sostener con fundamentos *ex post facto* que solo estarían limitados por la imaginación de los investigadores. ■

Como ilustra el Ejemplo 7.5.5, un investigador que utiliza una alternativa direccional cuando no está justificada paga el precio de doblar el riesgo de cometer un error de Tipo I. Es más, aquellos que leyeron el informe del investigador no se darían cuenta de que este riesgo se ha multiplicado por dos, y eso es por lo que algunos científicos recomiendan que nunca se utilicen alternativas direccionales.

Ejercicios 7.5.1-7.5.13

7.5.1 En cada uno de los siguientes conjuntos de datos utilice la Tabla 4 para acotar el P valor de una cola cuando dichos datos se analizan mediante un test t , suponiendo que la hipótesis alternativa es $H_A: \mu_1 > \mu_2$.

(a)

	Muestra 1	Muestra 2
n	10	10
$\bar{y}$	10,8	10,5
ET _($\bar{y}_1 - \bar{y}_2$) = 0,23 con $g_l = 18$		

(b)

	Muestra 1	Muestra 2
n	100	100
$\bar{y}$	750	730
$ET_{(\bar{y}_1 - \bar{y}_2)} = 11$ con $gl = 180$		

7.5.2 En cada uno de los siguientes conjuntos de datos, utilice la Tabla 4 para acotar el P valor de una cola cuando dichos datos se analizan mediante un test t , suponiendo que la hipótesis alternativa es $H_A: \mu_1 > \mu_2$.

(a)

	Muestra 1	Muestra 2
n	10	10
$\bar{y}$	3,24	3,00
$ET_{(\bar{y}_1 - \bar{y}_2)} = 0,61$ con $gl = 17$		

(b)

	Muestra 1	Muestra 2
n	6	5
$\bar{y}$	560	500
$ET_{(\bar{y}_1 - \bar{y}_2)} = 45$ con $gl = 8$		

(c)

	Muestra 1	Muestra 2
n	20	20
$\bar{y}$	73	79
$ET_{(\bar{y}_1 - \bar{y}_2)} = 2,8$ con $gl = 35$		

7.5.3 En cada una de las siguientes situaciones, suponga que se contrasta $H_0: \mu_1 = \mu_2$ frente a $H_A: \mu_1 > \mu_2$. Indique si hay o no evidencia significativa a favor de H_A .

- (a) $t_s = 3,75$ con 19 grados de libertad, $\alpha = 0,01$.
- (b) $t_s = 2,6$ con 5 grados de libertad, $\alpha = 0,10$.
- (c) $t_s = 2,1$ con 7 grados de libertad, $\alpha = 0,05$.
- (d) $t_s = 1,8$ con 7 grados de libertad, $\alpha = 0,05$.

7.5.4 En cada una de las siguientes situaciones, suponga que se contrasta $H_0: \mu_1 = \mu_2$ frente a $H_A: \mu_1 < \mu_2$. Indique si hay o no evidencia significativa a favor de H_A .

- (a) $t_s = -1,6$ con 23 grados de libertad, $\alpha = 0,05$.
- (b) $t_s = -2,3$ con 5 grados de libertad, $\alpha = 0,10$.
- (c) $t_s = 0,4$ con 16 grados de libertad, $\alpha = 0,10$.
- (d) $t_s = -2,8$ con 27 grados de libertad, $\alpha = 0,01$.

7.5.5 Unos investigadores en ecología midieron la concentración de glóbulos rojos en la sangre de 27 lagartos (*Sceloporus occidentalis*) capturados en el campo. Además, examinaron cada uno de los lagartos por si estaba infectado por el parásito de la malaria *Plasmodium*. Los valores del número de glóbulos rojos ($10^{-3} \times$ células por mm^3) fueron los que se indican en la tabla siguiente³⁵.

	Animales infectados	Animales no infectados
n	12	15
$\bar{y}$	972,1	843,4
s	245,1	251,2

Podría esperarse que la malaria redujera el número de glóbulos rojos, y de hecho en investigaciones previas con otras especies de lagartos se había encontrado ese efecto. ¿Dan los datos soporte a esta expectativa? Suponga que los datos están distribuidos normalmente. Contraste la hipótesis nula de que no existe diferencia contra la alternativa de que la población infectada tiene un número inferior de glóbulos rojos. Utilice un test t con

- (a) $\alpha = 0,05$
- (b) $\alpha = 0,10$

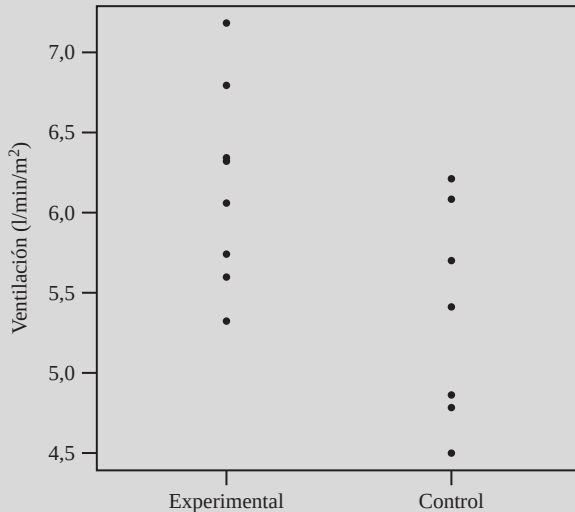
[Nota: la fórmula (6.71) da un valor de 24 gl].

7.5.6 Se emprendió un estudio para comparar las respuestas respiratorias de sujetos hipnotizados y no hipnotizados bajo ciertas instrucciones. Los 16 hombres voluntarios se asignaron aleatoriamente a un grupo experimental que iba a ser hipnotizado o a un grupo de control. Se tomaron medidas iniciales al comienzo del experimento. Al analizar los datos, los investigadores advirtieron que los patrones iniciales de respiración de los dos grupos eran diferentes. Esto fue sorprendente, ya que los sujetos habían sido tratados de la misma forma hasta ese momento. Una explicación propuesta para esta diferencia inesperada fue que el grupo experimental estaba más excitado anticipando la experiencia de ser hipnotizado. La tabla que sigue a continuación presenta un resumen de las medidas iniciales de ventilación total (litros de aire por minuto por metro cuadrado de área corporal). Los diagramas de puntos en paralelo de los datos se muestran en la gráfica siguiente³⁶. [Nota: la fórmula (6.7.1) da un valor de 14 gl].

	Experimental	Control
	5,32	4,50
	5,60	4,78
	5,74	4,79
	6,06	4,86
	6,32	5,41
	6,34	5,70
	6,79	6,08
	7,18	6,21
n	8	8
$\bar{y}$	6,169	5,291
s	0,621	0,652

- (a) Utilice un test t para contrastar la hipótesis de que hay diferencia contra una alternativa no direccional. Utilice $\alpha = 0,05$.

- (b) Utilice un test de t para contrastar la hipótesis de que no hay diferencia contra la alternativa de que las condiciones experimentales producen una media mayor que las condiciones de control. Utilice $\alpha = 0,05$.
- (c) ¿Cuál de los dos tests es más apropiado, el del apartado (a) o el del apartado (b)? Explique su respuesta.



7.5.7 En un estudio sobre el crecimiento de lechugas se cultivaron 10 tiestos que se asignaron aleatoriamente o bien a una solución de nutrientes estándar o a una solución que contenía nitrógeno extra. Después de 22 días de crecimiento, se recogieron las plantas y se pesaron, con los resultados que se muestran en la tabla siguiente³⁷. ¿Son suficientes los datos para concluir que el nitrógeno extra mejora el crecimiento de las plantas bajo estas condiciones? Utilice un test t con $\alpha = 0,05$ contra una alternativa direccional. (Suponga que los datos están distribuidos normalmente). [Nota: la fórmula (6.7.1) da un valor de 7,7 gl].

Solución de nutrientes	Peso de las hojas en seco (gm)		
	n	Media	DT
Estándar	5	3,62	0,54
Nitrógeno extra	5	4,17	0,67

7.5.8 La investigación ha demostrado que, en los mamíferos, dar a luz a un niño somete a las madres a más tensión que dar a luz una niña. ¿Afecta esto a la salud de su siguiente hijo? Un estudio comparó los pesos al nacer de los bebés después de haber tenido un niño y después de haber tenido una niña. La tabla que se presenta a continuación contiene los estadísticos resumen de una muestra de tamaño 76. Los datos parecen estar distribuidos normalmente³⁸. Utilice un test t , con $\alpha = 0,05$ y una alternativa direccional, para contrastar la hipótesis de la investigación de que el peso al nacer

es menor cuando el descendiente anterior es un niño. [Nota: la fórmula (6.7.1) da un resultado de 69,5 gl].

Sexo del descendiente anterior	Peso al nacer (kg)		
	n	Media	DT
Niño	33	3,32	0,62
Niña	43	3,63	0,63

7.5.9 Una entomóloga realizó un experimento para ver si dañar a una planta de tomate produciría cambios que mejoraran su defensa contra el ataque de los insectos. Hizo crecer las larvas del parásito del tabaco (*Manduca sexta*) en plantas dañadas y en plantas de control. La tabla que se presenta a continuación muestra los pesos (mg) de las larvas después de siete días de crecimiento³⁹. (Suponga que los datos están distribuidos normalmente). ¿Con que fuerza dan soporte los datos a las expectativas de la investigadora? Utilice un test t con un nivel de significación del 5 %. Sea H_A : que el daño a la planta tiende a disminuir el crecimiento de las larvas. [Nota: la fórmula (6.7.1) produce un resultado de 31,8 gl].

	Dañadas	Control
n	16	18
$\bar{y}$	28,66	37,96
s	9,02	11,14

7.5.10 Se probó la eficacia de una medicina contra el dolor en 50 mujeres que padecían dolores por calambres en el útero después de dar a luz. Se asignaron aleatoriamente 25 mujeres al grupo que recibió la medicina y las restantes 25 recibieron un placebo (sustancia inerte). Se suministraron cápsulas de la medicina o del placebo antes del desayuno y al mediodía. Se calculó para cada mujer una puntuación de alivio del dolor, basada en preguntas realizadas cada hora a lo largo del día. Los posibles valores de las puntuaciones de alivio del dolor variaban desde cero (ningún alivio) hasta 56 (alivio completo durante ocho horas). La tabla que sigue a continuación presenta el resumen de los resultados⁴⁰. [Nota: la fórmula (6.7.1) da un resultado de 47,2 gl].

Tratamiento	Puntuación de alivio del dolor		
	n	Media	DT
Medicina	25	31,96	12,05
Placebo	25	25,32	13,78

- (a) Realice un contraste sobre la evidencia de eficacia utilizando un test t . Utilice una alternativa direccional y $\alpha = 0,05$.
- (b) Si la hipótesis alternativa fuera no direccional, ¿cómo cambiaría su respuesta al apartado (a)?

7.5.11 El íleo postoperatorio (IPO) es una forma de disfunción gastrointestinal que aparece comúnmente después de una cirugía abdominal y tiene como resultado una movilidad gastrointestinal ausente o retrasada. ¿Reduce el íleo postoperatorio (IPO) el hecho de mecerse en una silla después de recibir cirugía abdominal)? Se dividieron 66 pacientes de postoperatorio de una cirugía abdominal aleatoriamente en dos grupos. El grupo experimental ($n = 34$) recibió cuidados estándar junto con el uso de una mecedora mientras que el grupo de control ($n = 32$) recibió solo los cuidados estándar. En cada paciente, se midió el tiempo de postoperatorio hasta el primer flato (días) (una indicación de que el IPO ha terminado). Los resultados se muestran a continuación⁴¹.

	Tiempo hasta el primer flato		
	n	Media (días)	DT
Mecedora	34	3,16	0,86
Control	32	3,88	0,80

- (a) ¿Existe evidencia de que el uso de la mecedora reduce la duración del IPO (es decir, el tiempo hasta el primer flato)? Utilice un test t con un alternativa direccional y $\alpha = 0,05$.
- (b) Aunque los investigadores hicieron la hipótesis de que el uso de una mecedora podría reducir la duración del IPO, es también razonable realizar la hipótesis de que el uso de una mecedora podría aumentar la duración del IPO. Basándose en esta posibilidad, comente si es apropiado el uso de un test direccional o de un test no direccional. (Sugerencia: considere lo que las recomendaciones médicas podrían haber hecho basándose en esta investigación).

7.5.12 En el Ejemplo 7.2.6 consideramos el contraste de $H_0: \mu_1 = \mu_2$ frente a la hipótesis alternativa no direccional $H_A: \mu_1 \neq \mu_2$ y obtuvimos que el P valor se podía acotar como $0,06 < P \text{ valor} < 0,10$. Recuerdese que la media muestral del grupo 1 (el grupo de control) era de 15,9, que era menor que la media muestral de 11,0 del grupo 2 (el grupo tratado con Ancimidol). Sin embargo, el Ancimidol se considera un inhibidor del crecimiento, lo que significa que se podría esperar que el grupo de control tuviera una media mayor que el grupo de tratamiento si el Ancimidol tuviera algún efecto sobre el tipo de planta que estaba siendo estudiada (en este caso, la Planta Rápida de Wisconsin). Suponga que el investigador tenía la expectativa de que el Ancimidol retardaría al crecimiento (antes de realizar el experimento) y que hubiera realizado un contraste de $H_0: \mu_1 = \mu_2$ frente a la hipótesis alternativa no direccional $H_A: \mu_1 > \mu_2$, utilizando $\alpha = 0,05$. ¿Cuáles serían los límites en el P valor? ¿Se rechazaría H_0 ? ¿Por qué o por qué no? ¿Cuál sería la conclusión del experimento? (Nota: este problema prácticamente no requiere cálculos).

zaría H_0 ? ¿Por qué o por qué no? ¿Cuál sería la conclusión del experimento? (Nota: este problema prácticamente no requiere cálculos).

7.5.13 (Ejercicio para computador) Un ecólogo estudió el hábitat de un pez de arrecife marino, el pez verde de seis barras (*Thalassoma hardwicke*), cerca de una isla en la Polinesia Francesa rodeada de una barrera de arrecife. Examinó 48 parcelas de asentamiento en el arrecife en dos distancias medidas desde la cresta del arrecife: 250 metros desde la cresta y 800 metros desde la cresta. En cada parcela del arrecife calculó la «densidad de pobladores» es decir, el número de pobladores (peces jóvenes) por unidad de hábitat de asentamiento. Antes de tomar los datos, realizó la hipótesis de que la densidad de pobladores decrecería con la distancia medida desde la cresta del arrecife, ya que la forma en la que las olas rompen contra la cresta del arrecife produce recursos (por ejemplo, alimento) que tienden a disminuir con la distancia a la cresta del arrecife. Se presentan a continuación los datos⁴²:

250 metros			800 metros		
0,318	0,758	0,318	0,941	0,289	0,399
0,637	0,372	0,524	0,279	0,392	0,955
0,196	0,637	1,404	1,021	0,725	0,531
0,624	1,560	0,000	0,108	1,318	0,252
0,909	0,207	1,061	0,738	0,612	1,179
0,295	0,685	0,590	0,907	0,637	0,442
0,594	0,000	0,363	0,503	0,181	0,291
0,442	1,303	1,567	0,637	0,941	0,579
1,220	0,898	1,577	1,498	0,265	0,252
1,303	1,157	0,312	0,866	0,979	0,373
0,187	0,970	0,758	0,588	0,909	0,000
1,560	0,624	0,505	0,606	0,283	0,463
0,849	1,592	0,909	0,490	0,337	1,248
2,411	1,019	0,362	0,163	0,813	2,010
1,705	0,829	0,329	0,277	0,000	1,213
1,019	0,884	0,909	0,293	0,544	0,808

Para 250 metros, la media muestral es 0,818 y la DT muestral es 0,514. Para 800 metros, la media muestral es 0,628 y la DT muestral es 0,413. ¿Proporcionan estos datos evidencia estadística significativa, con un nivel de significación de 0,10, para dar soporte a la teoría del ecólogo? Investíguelo con una gráfica y un contraste adecuados.

7.6 Más sobre la interpretación de la significación estadística

Idealmente, el análisis estadístico debería ser de utilidad al investigador ayudándole a clarificar qué mensaje está contenido en los datos. Para este propósito, no es suficiente que los cálculos estadísticos sean correctos. Los resultados

deben ser también interpretados correctamente. En esta sección exploraremos algunos principios de interpretación que se aplican no solamente al test t , sino también a otros tests estadísticos que se presentarán posteriormente.

Diferencia significativa frente a diferencia importante

El término *significativo* se utiliza a menudo para describir los resultados de un análisis estadístico. Por ejemplo, si el experimento que compara un medicamento contra un placebo produce datos con un P valor muy pequeño, entonces la conclusión se podría expresar como «el efecto del medicamento fue altamente significativo». Como otro ejemplo, si dos fertilizantes para el trigo resultan en una comparación de las cosechas con un P valor grande, entonces la conclusión se podría expresar como «las cosechas de trigo no difieren significativamente entre los dos fertilizantes» o «la diferencia entre los fertilizantes no fue significativa». Como un tercer ejemplo, supongamos que se comprueban los efectos tóxicos de una sustancia comparando animales expuestos y animales de control, y que no se rechaza la hipótesis nula de no diferencia. Entonces la conclusión podría expresarse como «no se ha encontrado toxicidad significativa».

Resulta claro que esta fraseología que utiliza el término *significativo* puede resultar seriamente engañosa. Después de todo, en el uso del castellano, la palabra significativo tiene la connotación de «sustancial» o «importante». En la jerga estadística, sin embargo, la frase

«La diferencia fue significativa»

no significa ni más ni menos que

«La hipótesis nula de no diferencia se rechaza».

Es decir, «hemos encontrado suficiente evidencia de que la diferencia en las medias muestrales no está causada solo por el error debido al azar».

Por el mismo motivo, la frase

«La diferencia no fue significativa»

significa

«No hubo evidencia suficiente de que la diferencia observada en las medias fuera debida a otra cosa que a la variación causada por el azar».

Quizá sería preferible que se hubiera utilizado otra palabra diferente en lugar de «significativo», como «discernible» (queriendo decir que el test t discierne una diferencia). Lamentablemente, el uso especializado de la palabra *significativo* se ha hecho muy común en la literatura científica e incomprensiblemente es la fuente de mucha confusión.

Es esencial reconocer que un test estadístico proporciona información solo sobre una cuestión: ¿Es la diferencia observada en los datos lo suficientemente grande como para inferir que existe en la población una diferencia en la misma dirección? La cuestión de si una diferencia es importante, en contraposición a (estadísticamente) significativa, no se puede decidir solo con la base de los P valores, sino que debe incluir también un examen de la magnitud de la diferencia estimada en las poblaciones, así como la experiencia específica en el área de investigación o en la situación práctica. Los dos ejemplos siguientes ilustran este hecho.

Ejemplo 7.6.1 LD en suero

La lactato deshidrogenasa (LD) es una enzima que puede mostrar una actividad elevada después de que el músculo cardíaco u otros tejidos han sido dañados. Un estudio grande sobre los niveles de LD en suero en personas jóvenes saludables produjo los resultados que se muestran en la Tabla 7.6.1⁴³.

Tabla 7.6.1 LD en suero		
	Hombres	Mujeres
n	270	264
$\bar{y}$	60	57
s	11	10

La diferencia entre hombres y mujeres es bastante significativa: de hecho, $t_s = 3,3$, que da un P valor $\approx 0,001$. Sin embargo, esto no implica que la diferencia ($60 - 57 = 3$ U/l) sea grande o importante en ningún sentido práctico. ■

Ejemplo 7.6.2 Peso corporal

Imaginemos que estamos estudiando el peso corporal de hombres y mujeres, y que tenemos los datos ficticios pero realistas que se muestran en la Tabla 7.6.2⁴⁴.

Tabla 7.6.2 Peso corporal		
	Hombres	Mujeres
n	2	2
$\bar{y}$	175	143
s	35	34

Para estos datos, el test t da un valor $t_s = 0,93$ y un P valor $\approx 0,45$. La diferencia observada entre hombres y mujeres no es pequeña (es $175 - 143 = 32$ lb), aunque no es estadísticamente significativa para ninguna elección razonable de α . La falta de significación estadística no implica que la diferencia de peso corporal entre los sexos sea pequeña o no sea importante. Sólo significa que los datos son inadecuados para caracterizar dicha diferencia en las medias poblacionales. Una diferencia muestral de 32 lb podría suceder muy fácilmente por azar si las dos poblaciones fueran idénticas, especialmente con tamaños de las muestras tan pequeños. ■

Tamaño del efecto

Los ejemplos anteriores muestran que la significación o no significación de una diferencia no indica si la diferencia es importante. No obstante, la cuestión de la «importancia» puede y debería ser tomada en cuenta en la mayor parte de los análisis de datos. Para evaluar la importancia, es necesario considerar la *magnitud* de la diferencia. En el Ejemplo 7.6.1 la diferencia de hombres frente a mujeres es «estadísticamente significativa», pero esto es debido en gran parte a que los tamaños de las muestras son bastante grandes. El test t utiliza el estadístico de contraste

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2)}{ET_{(\bar{y}_1 - \bar{y}_2)}}$$

Si n_1 y n_2 son grandes, entonces $ET_{(\bar{y}_1 - \bar{y}_2)}$ será pequeña y el estadístico de contraste tenderá a ser grande incluso aunque la diferencia en las medias observadas ($\bar{Y}_1 - \bar{Y}_2$) sea muy pequeña. Por tanto, podría encontrarse evidencia significativa a favor de H_A debido a que el tamaño de la muestra es grande, incluso si μ_1 y μ_2 son casi iguales. El tamaño de la muestra se comporta como una lente de aumento: *cuanto mayor sea el tamaño de la muestra, menor es la diferencia que se puede detectar en un contraste de hipótesis*.

El **tamaño del efecto** en un estudio es la diferencia entre μ_1 y μ_2 , expresada en relación a la desviación típica de una de las poblaciones. Si las dos poblaciones tienen la misma desviación típica, σ , entonces el tamaño del efecto es*:

$$\text{Tamaño del efecto} = \frac{|\mu_1 - \mu_2|}{\sigma}$$

Por supuesto, cuando se trabaja con datos de una muestra solo podemos calcular el tamaño del efecto *estimado* utilizando los valores muestrales en lugar de los valores poblacionales desconocidos.

Ejemplo 7.6.3 LD en suero

En los datos dados en el Ejemplo 7.6.1, la diferencia de las medias muestrales, $60 - 57 = 3$, es menor que la tercera parte de una desviación típica. Utilizando la mayor DT muestral podemos calcular un tamaño del efecto muestral de

$$\text{Tamaño del efecto} = \frac{|\bar{y}_1 - \bar{y}_2|}{s} = \frac{60 - 57}{11} = 0,27$$

* Si las desviaciones típicas no son iguales, podemos utilizar la mayor DT para definir el tamaño del efecto.

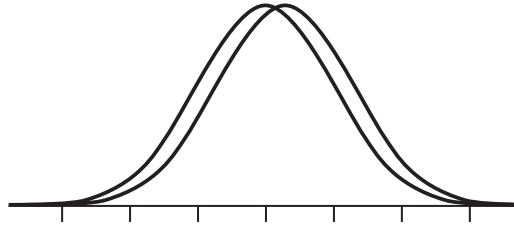


Figura 7.6.1 Solapamiento entre dos poblaciones distribuidas normalmente cuando el tamaño del efecto es 0,27

Esto indica que existe mucho solapamiento entre los dos grupos. La Figura 7.6.1 muestra la extensión del solapamiento que aparece entre dos poblaciones distribuidas normalmente que se diferencian en promedio en 0,27 DT. ■

Ejemplo 7.6.4 Peso corporal

En los datos dados en el Ejemplo 7.6.2, la diferencia en las medias muestrales, $175 - 143 = 32$, es aproximadamente una desviación típica. El tamaño del efecto muestral es

$$\text{Tamaño del efecto} = \frac{|\bar{y}_1 - \bar{y}_2|}{s} = \frac{175 - 143}{35} = 0,91$$

La Figura 7.6.2 muestra la extensión del solapamiento que aparece si dos poblaciones distribuidas normalmente se diferencian en promedio en 0,91 DT. ■

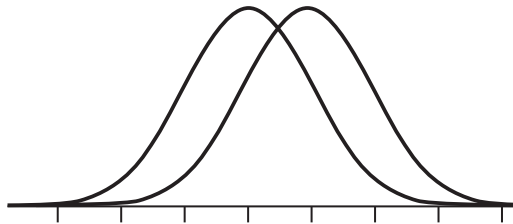


Figura 7.6.2 Solapamiento entre dos poblaciones distribuidas normalmente cuando el tamaño del efecto es 0,91

La definición de tamaño del efecto que estamos utilizando probablemente no resulte familiar al lector orientado a la biología. Es más común en biología «estandarizar» una diferencia de dos cantidades expresándola como un porcentaje de una de ellas. Por ejemplo, la diferencia de pesos dados en la Tabla 7.6.2 entre hombres y mujeres, expresada como porcentaje del peso medio de las mujeres, es

$$\frac{\bar{y}_1 - \bar{y}_2}{\bar{y}_2} = \frac{175 - 143}{143} = 0,22 \text{ o } 22 \%$$

Por tanto, los hombres son aproximadamente un 22 % más pesados que las mujeres. Sin embargo, desde un punto de vista estadístico, a menudo es más relevante decir que los pesos medios de los hombres y las mujeres se diferencian 0,91 DT.

Intervalos de confianza para evaluar la importancia

Calcular el tamaño del efecto es una forma de cuantificar cómo están de separadas dos medias muestrales. Otro planteamiento razonable es utilizar la diferencia observada ($\bar{Y}_1 - \bar{Y}_2$) para construir un intervalo de confianza para la diferencia poblacional ($\mu_1 - \mu_2$). Al interpretar el intervalo de confianza, el juicio de lo que es «importante» se realiza sobre la base de la experiencia en una situación práctica concreta. Los tres ejemplos siguientes ilustran este uso de los intervalos de confianza.

Ejemplo 7.6.5 LD en suero

En los datos de LD del Ejemplo 7.6.1, un intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ es

$$3 \pm 1,8$$

o

$$(1,2, 4,8)$$

Este intervalo implica (con un 95 % de confianza) que la diferencia poblacional de las medias entre los sexos no excede las 4,8 U/l. Como experto, un médico que evalúa esta información podría saber que la fluctuación típica de un día a otro en el nivel de LD de una persona es aproximadamente 6,5 U/l, que es mayor que 4,8 U/l, el valor más alto estimado para la diferencia de las medias entre los sexos. Y, por lo tanto, esta diferencia es despreciable desde un punto de vista médico. En consecuencia, el médico podría concluir que no es necesario diferenciar entre los sexos al establecer umbrales clínicos para el diagnóstico de una enfermedad. En este caso, se diría que la diferencia de LD entre los sexos es estadísticamente significativa pero no es médicamente importante. Expresando esto de otra forma, los datos sugieren que los hombres de hecho tienden a tener niveles más altos que las mujeres, pero no tan altos como para que tengan alguna forma de utilidad clínica. ■

Ejemplo 7.6.6 Peso corporal

En los datos sobre el peso corporal del Ejemplo 7.6.2, un intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$ es

$$32 \pm 149$$

o

$$(-117, 181)$$

A partir de este intervalo de confianza no podemos decidir si la verdadera diferencia (entre las medias poblacionales) es grande favoreciendo a las mujeres, es pequeña, o es grande favoreciendo a los hombres. Debido a que el intervalo de confianza contiene números de magnitud pequeña y grande, no nos dice si la diferencia entre los sexos es importante o no es importante. Con un intervalo de confianza tan grande, un investigador probablemente desearía realizar un estudio mayor para evaluar mejor la importancia de la diferencia. Supongamos, por ejemplo, que las medias y las desviaciones típicas fueran como las que se dan en la Tabla 7.6.2, pero estuvieran basadas en 2.000 personas en vez de en 2 personas de cada sexo. Entonces el intervalo de confianza del 95 % sería

$$32 \pm 2$$

o

$$(30, 34)$$

Este intervalo implicaría (con un 95 % de confianza) que la diferencia es como mínimo de 30 lb, una cantidad que podría ser razonablemente considerada como importante, al menos para algunos propósitos. ■

Ejemplo 7.6.7 Cosecha de tomates

Supongamos que un horticultor está comparando las cosechas de dos variedades de tomates. Las cosechas se miden en libras de tomates por planta. Sobre la base de consideraciones prácticas, el horticultor ha decidido que una diferencia de las variedades es «importante» solo si supera una libra por planta, en promedio. Es decir, la diferencia es importante si

$$|\mu_1 - \mu_2| > 1,0 \text{ lb}$$

Supongamos que los datos del horticultor proporcionan el siguiente intervalo de confianza del 95 %:

$$(0,2, 0,3)$$

Como la mayor estimación para la diferencia de las poblaciones es solo 0,3 libras (todos los valores del intervalo son menores que 1,0 libras), los datos dan soporte (con una confianza del 95 %) a la aseveración de que la diferencia no es importante de acuerdo con el criterio del horticultor. ■

En muchas investigaciones, son de interés tanto la significación estadística como la importancia práctica. El siguiente ejemplo ilustra cómo se puede visualizar la relación entre estos dos conceptos utilizando intervalos de confianza.

Ejemplo 7.6.8 Cosecha de tomates

Volvamos al experimento sobre tomates del Ejemplo 7.6.7. El intervalo de confianza era

$$(0,2, 0,3)$$

Recordemos de la Sección 7.3 que el intervalo de confianza se puede interpretar en función de un test t . Como todos los valores dentro del intervalo de confianza son positivos, un test t (de dos colas) con $\alpha = 0,05$ encontrará evidencia significativa a favor de H_A . Por tanto, la diferencia entre las dos variedades es estadísticamente significativa, aunque no es importante desde el punto de vista de la horticultura. Los datos indican que la variedad 1 es mejor que la variedad 2, pero también que no es mucho mejor. La distinción entre significación e importancia en este ejemplo se puede ver en la Figura 7.6.3, donde se muestra el intervalo de confianza situado en un eje $(\mu_1 - \mu_2)$. Nótese que el intervalo de confianza está completamente a un lado del cero y también completamente a un lado del umbral de «importancia» de 1,0.

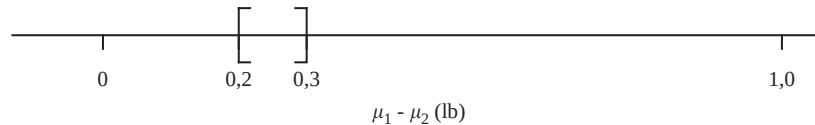


Figura 7.6.3 Intervalo de confianza en el Ejemplo 7.6.8

Para profundizar más en la relación entre significación e importancia, consideremos otros posibles resultados del experimento de los tomates. La Tabla 7.6.3 muestra cómo interpretaría el horticultor varios posibles intervalos de confianza, utilizando el criterio de que para que una diferencia se considere importante debe superar el valor de 1,0 libras.

Tabla 7.6.3 Interpretación de intervalos de confianza		
Intervalo de confianza del 95 %	¿Es la diferencia	
	significativa?	importante?
(0,2, 0,3)	Sí	No
(1,2, 1,3)	Sí	Sí
(0,2, 1,3)	Sí	No se puede decir
(-0,2, 0,3)	No	No
(-1,2, 1,3)	No	No se puede decir

La Tabla 7.6.3 muestra que una diferencia significativa puede ser o no puede ser importante, y que una diferencia importante puede ser o no puede ser significativa. En la práctica, la valoración de la importancia utilizando intervalos de confianza es un complemento simple y extremadamente útil a un contraste de hipótesis. ■

Ejercicios 7.6.1-7.6.8

7.6.1 Se realizó un experimento de campo para evaluar un nuevo tratamiento de semillas que se suponía mejoraba la cosecha de soja. Cuando un estadístico analizó los datos, encontró que el rendimiento medio de las semillas tratadas era de 40 lb/acre mayor que las de las parcelas de control donde

se han plantado semillas no tratadas. Sin embargo, el estadístico declaró que la diferencia «no era (estadísticamente) significativa». Los proponentes del tratamiento pusieron serias objeciones a la afirmación del estadístico, señalando que, a los precios actuales del mercado, 40 lb/acre, supondría una

suma considerable, que sería altamente significativa para el granjero. ¿Cómo respondería a esta objeción?⁴⁵

7.6.2 En un estudio clínico sobre tratamientos de la artritis reumatoide se asignó aleatoriamente a una serie de pacientes para recibir o bien una medicación estándar o una medicación de nuevo diseño. Después de un periodo observación adecuado, el análisis estadístico demostró que no había diferencias significativas en la respuesta terapéutica de los dos grupos, pero que la incidencia de efectos colaterales indeseables era significativamente menor en el grupo que recibió la nueva medicación. Los investigadores concluyeron que la nueva medicación debería ser considerada como claramente preferible a la medicación estándar, porque había demostrado ser igualmente efectiva terapéuticamente y producía menos efectos colaterales. ¿En qué aspecto falla el razonamiento de los investigadores? (Suponga que el término «significativo» se refiere al rechazo de H_0 con $\alpha = 0,05$).

7.6.3 Existe una antigua creencia popular sobre que el sexo de un bebé se puede adivinar antes de su nacimiento basándose en su frecuencia cardíaca. En una investigación para comprobar esta teoría, se observaron frecuencias cardíacas fetales en madres admitidas a una sala de maternidad. Los resultados (en latidos por minuto) se resumen en la tabla siguiente⁴⁶.

	Frecuencia cardíaca (latidos por minuto)		
	n	Media	ET
Hombres	250	137,21	0,62
Mujeres	250	137,18	0,53

Construya un intervalo de confianza del 95 % para la diferencia de las medias poblacionales. ¿Da soporte el intervalo de confianza a la afirmación de que la diferencia de las medias poblacionales de las frecuencias cardíacas de los sexos (si existe) es pequeña y no es importante? (Utilice su propio conocimiento «de experto» sobre la frecuencia cardíaca para realizar un juicio de lo que «no es importante»).

7.6.4 El ácido comárico es un componente que tiene un papel en la resistencia del maíz a las enfermedades. Un botánico midió la concentración de ácido comárico en las semillas de maíz que habían crecido en la oscuridad o con un fotoperiodo de luz/oscuridad. Los resultados (en nmol de ácido por gramo de tejido) se muestran en la tabla siguiente⁴⁷. [Nota: la fórmula (6.7.1) a un valor de 5,7 gl].

Suponga que un botánico considera que el efecto de las condiciones de iluminación es «importante» si la diferencia de medias es del 20 %, es decir, aproximadamente 20 nmol/g.

Basándose en un intervalo de confianza del 95 %, ¿indican los datos que la verdadera diferencia es «importante»?

	Oscuridad	Fotoperiodo
n	4	4
$\bar{y}$	106	102
s	21	27

7.6.5 Repita el Ejercicio 7.6.4, suponiendo que las medias y las desviaciones típicas son las que se muestran en la tabla, pero que los tamaños muestrales son 10 veces mayores (es decir, $n = 40$ para la «oscuridad» y $n = 40$ para el «fotoperiodo»). [Nota: la fórmula (6.7.1) da un resultado de 73,5 gl].

7.6.6 Unos investigadores midieron las alturas, en mm, de los tobillos de 460 jóvenes (de edades entre 11 y 16 años). Los resultados se muestran en la tabla siguiente⁴⁸.

	Hombres	Mujeres
n	244	216
$\bar{y}$	55,3	53,3
s	6,1	5,4

Calcule el tamaño del efecto muestral a partir de estos datos.

7.6.7 Como parte de un amplio estudio sobre la química del suero en personas saludables, se obtuvieron los siguientes datos sobre concentración en suero de ácido úrico en hombres y mujeres entre 18 y 55 años de edad⁴⁹.

	Ácido úrico en suero (mmol/l)	
	Hombres	Mujeres
n	530	420
$\bar{y}$	0,354	0,263
s	0,058	0,051

Obtenga un intervalo de confianza de 95 % para la verdadera diferencia de las medias poblacionales. Suponga que los investigadores piensan que la diferencia de las medias poblacionales es «clínicamente importante» si supera los 0,08 nmol/l. ¿Indica el intervalo de confianza si la diferencia es «clínicamente importante»? [Nota: la fórmula (6.7.1) da un resultado de 934 gl].

7.6.8 Repita el Ejercicio 7.6.7, suponiendo que las medias y las desviaciones típicas son las que se dan en la tabla, pero los tamaños muestrales son solamente la décima parte (es decir, 53 hombres y 42 mujeres). [Nota: la fórmula (6.7.1) da un resultado de 92 gl].

7.7 Planificación de la potencia adecuada (opcional)

La definición de la potencia de un contraste estadístico es

$$\text{Potencia} = \Pr\{\text{evidencia significativa a favor de } H_A\} \text{ si } H_A \text{ es cierta}$$

Expresándolo de otra forma, la potencia de un contraste es la probabilidad de obtener datos que proporcionan evidencia estadísticamente significativa a favor de H_A cuando H_A es cierta.

Como la potencia es la probabilidad de no cometer un error (de Tipo II), es deseable una potencia alta. Si H_A es cierta, el investigador desearía descubrirlo al realizar su estudio. Pero la potencia tiene un precio. A igualdad de todos los otros factores, cuanto más observaciones (muestras más grandes) más potencia, pero las observaciones cuestan tiempo y dinero. En esta sección explicaremos cómo pueden los investigadores planificar racionalmente sus experimentos para obtener una potencia adecuada a efectos de su proyecto de investigación, con el coste tan pequeño como sea posible.

Concretamente, consideraremos la potencia de un test t de dos muestras, realizado con un nivel de significación α . Supondremos que las poblaciones son normales con DT iguales y denominaremos σ al valor común de la DT (es decir, $\sigma_1 = \sigma_2 = \sigma$). Se puede demostrar que, en este caso, para un tamaño total de la muestra de $2n$, la potencia se maximiza si los tamaños muestrales son iguales. Por tanto, supondremos que n_1 y n_2 son iguales y denominaremos n a ese valor común (es decir, $n_1 = n_2 = n$).

Bajo las condiciones anteriores, la potencia del test t depende de los siguientes factores (a) α ; (b) σ ; (c) n , y (d) $(\mu_1 - \mu_2)$. Tras discutir brevemente cada uno de estos factores, los centraremos en la importante cuestión de escoger el valor de n .

Dependencia de la potencia con α

Al escoger α , se elige un nivel de protección contra los errores de Tipo I. Sin embargo, esta protección tiene un compromiso con la vulnerabilidad a los errores de Tipo II. Si, por ejemplo, se escoge $\alpha = 0,01$ en vez de $\alpha = 0,05$, entonces se requiere una evidencia más fuerte a favor de H_A antes de afirmar que existe evidencia significativa a favor de H_A y, por tanto (quizá sin ser consciente de ello) se está aumentando el riesgo de cometer un error de Tipo II y reduciendo la potencia. Por tanto, hay un compromiso inevitable entre el riesgo de cometer un error de Tipo I y el riesgo de cometer un error de Tipo II.

Dependencia con σ

Cuanto mayor sea σ , menor será la potencia (si todos los otros factores son iguales). Recuérdese del Capítulo 5 que la fiabilidad de la media muestral está determinada por el valor

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

Cuanto mayor sea σ , mayor variabilidad tendrá la media muestral. Por tanto, tener un valor de σ mayor implica tener muestras que producen información menos fiable sobre cada media poblacional y, por tanto, menos potencia para discernir una diferencia entre ellas. Entonces, para aumentar la potencia, los investigadores generalmente intentan diseñar su investigación de forma que σ sea tan pequeña como sea posible. Por ejemplo, un botánico intentaría mantener condiciones de iluminación constante en un área de invernadero, un farmacólogo intentará utilizar animales de experimentación genéticamente idénticos, etc. Generalmente, sin embargo, σ no se puede reducir a cero. Puede existir todavía una variación considerable en las observaciones.

Dependencia con n

Cuanto mayor sea n , mayor será la potencia (si todos los otros factores se mantienen iguales). Si aumentamos n , disminuimos $\sigma/\sqrt{n}$. Esto hace que mejore la precisión de las medias muestrales ($\bar{Y}_1$ e $\bar{Y}_2$). Además, valores mayores de n dan más información sobre σ . Esto se refleja en un menor valor crítico del test (que se reduce porque hay más grados de libertad). Por tanto, aumentar n aumenta la potencia del test de dos formas.

Dependencia con $(\mu_1 - \mu_2)$

Además de los factores que hemos comentado, la potencia del test t depende también de la diferencia real entre las medias poblacionales, es decir, depende de $(\mu_1 - \mu_2)$. Esta dependencia es muy natural, como se ilustra en el siguiente ejemplo.

Ejemplo 7.7.1 Alturas de personas

Para ilustrar claramente los conceptos, consideraremos una variable familiar, la altura corporal de las personas. Imaginemos lo que sucedería si un investigador midiera las alturas de dos muestras aleatorias de 11 personas cada una ($n = 11$), y después realizara un test t de dos colas con $\alpha = 0,05$.

- En primer lugar, supongamos que la muestra 1 está formada por hombres de 17 años de edad y la muestra 2 por mujeres de 17 años de edad. Las dos medias poblacionales difieren sustancialmente; de hecho, $(\mu_1 - \mu_2)$ es aproximadamente 5 pulgadas ($\mu_1 \approx 69,1$ y $\mu_2 \approx 64,1$ pulgadas)⁵⁰. Se puede demostrar (como veremos) que, en este caso, el investigador tiene aproximadamente el 99 % de probabilidad de obtener evidencia significativa a favor de una diferencia (es decir, H_A) y concluir correctamente que los hombres en la población de 17 años de edad son más altos (en promedio) que las mujeres.
- Por el contrario, supongamos que la muestra 1 consiste en mujeres de 17 años de edad y la muestra 2 consiste en mujeres de 14 años de edad. Las dos medias poblacionales son diferentes, pero solo en una cantidad modesta. La diferencia es $(\mu_1 - \mu_2) = 0,6$ pulgadas ($\mu_1 \approx 64,1$ y $\mu_2 \approx 63,5$). Se puede demostrar que en este caso el investigador tiene menos de un 10 % de probabilidad de obtener evidencia significativa a favor de una diferencia (es decir, H_A). En otras palabras, hay más de un 90 % de probabilidad de que el investigador no detecte el hecho de que las chicas de 17 años de edad son más altas que las de 14 años de edad. (De hecho, se puede demostrar que hay un 29 % de probabilidad de que $\bar{Y}_1$ sea menor que $\bar{Y}_2$). Es decir, ¡hay un 29 % de probabilidad de que 11 chicas de 17 años de edad escogidas aleatoriamente sean en promedio más bajas que 11 chicas de 14 años de edad escogidas aleatoriamente!).

La diferencia entre los casos (a) y (b) no es debida a ningún cambio en las DT. De hecho, en cada una de las tres poblaciones el valor de σ es aproximadamente 2,5 pulgadas. Más bien, la diferencia es debida al hecho simple de que, con un valor fijo de n y σ , es más fácil detectar una diferencia grande que una diferencia pequeña. ■

Planificación de un estudio

Supongamos que un investigador está planificando un estudio en el cual el uso de un test t es apropiado. ¿Cómo tendrá en cuenta todos los factores que influyen en la potencia del contraste? Consideremos primero la selección del nivel de significación α . Un planteamiento simple es empezar por determinar el coste de un estudio con potencia adecuada utilizando una elección por exceso de α (es decir, $\alpha = 0,05$ o $0,10$). Si el coste no es alto, el investigador puede considerar la reducción de α (por ejemplo, a $0,01$) y ver si todavía se puede llevar a cabo un estudio con la potencia adecuada.

Supongamos, entonces, que el investigador ha elegido un valor de trabajo de α . Supongamos también que el experimento se ha diseñado para reducir σ tanto como sea posible y que el investigador tiene disponible una estimación inicial del valor de σ .

En este punto, el investigador necesita preguntarse sobre la magnitud de la diferencia que desea detectar. Como vimos en el Ejemplo 7.7.1, un tamaño de la muestra dado puede ser adecuado para detectar una diferencia grande en las medias poblacionales, pero absolutamente inadecuado para detectar una diferencia pequeña. Como un ejemplo más realista, un experimento que utilice 5 ratas en un grupo de tratamiento y 5 ratas en un grupo de control podría tener un tamaño de muestra suficiente para detectar un efecto sustancial del tratamiento, pero la detección de un efecto sutil del tratamiento requeriría más ratas (quizá 30) en cada grupo.

La presentación anterior sugiere que escoger un tamaño de muestra para obtener una potencia adecuada es algo análogo a escoger un microscopio: necesitamos alta potencia de resolución si queremos ver estructuras muy finas; para estructuras más grandes puede servir una lente más normal. Para seguir con la planificación del experimento, el investigador necesita decidir el tamaño del efecto que está buscando.

Recuérdese que en la Sección 7.7 definimos el tamaño del efecto de un estudio como la diferencia entre μ_1 y μ_2 , expresada en relación a la desviación típica de una de las poblaciones. Si como estamos suponiendo, las dos poblaciones tienen la misma desviación típica, σ , entonces el tamaño del efecto es

$$\text{Tamaño del efecto} = \frac{|\mu_1 - \mu_2|}{\sigma}$$

Es decir, el tamaño del efecto es la diferencia en las medias poblacionales expresada en relación a la DT poblacional común. El tamaño del efecto es un tipo de «relación señal y ruido», donde $(\mu_1 - \mu_2)$ representa la señal que deseamos detectar y σ representa el ruido de fondo que tiende a oscurecer la señal. La Figura 7.7.1(a) muestra dos curvas normales en las cuales el tamaño del efecto es 0,5. La Figura 7.7.1(b) muestra dos curvas normales en las cuales el tamaño del efecto es 4. Claramente, dado un tamaño de muestra fijo, es más fácil detectar la diferencia entre las curvas en la gráfica (b) que en la gráfica (a).

Si α y el tamaño del efecto se han especificado, entonces la potencia del test t depende solo de los tamaños de la muestra (n). La Tabla 5 al final del libro muestra el valor de n requerido para obtener una potencia especificada en función de un tamaño del efecto especificado. Veamos cómo se aplica la Tabla 5 a nuestro ejemplo familiar de la altura corporal.

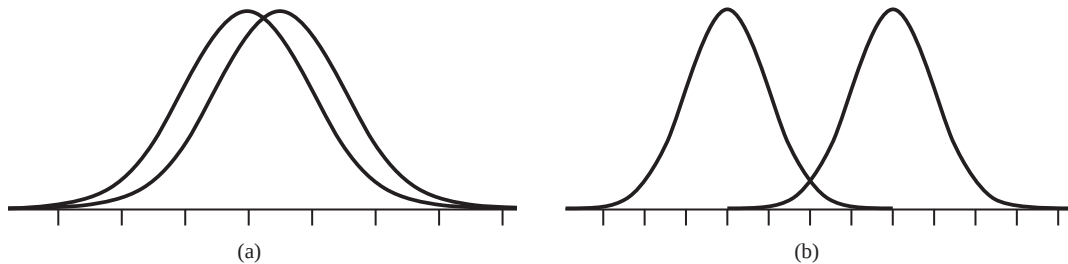


Figura 7.7.1 Distribuciones normales con el tamaño del efecto (a) de 0,5 y (b) de 4

Ejemplo 7.7.2 Altura de personas

En el Ejemplo 7.7.1, caso (a), consideramos muestras de hombres y mujeres de 17 años de edad. El tamaño del efecto es

$$\frac{|\mu_1 - \mu_2|}{\sigma} = \frac{|69,1 - 64,1|}{2,5} = \frac{5}{2,5} = 2,0$$

En el caso de un test t de dos colas con $\alpha = 0,05$, la Tabla 5 muestra que el tamaño de la muestra necesario para tener una potencia de 0,99 es $n = 11$. Esta es la base para la afirmación del Ejemplo 7.7.1 de que el investigador tiene un 99 % de probabilidad de detectar la diferencia entre hombres y mujeres. La Figura 7.7.2 muestra las dos distribuciones consideradas en el Ejemplo 7.7.2. Supongamos que 100 investigadores realizan el siguiente estudio. Tomar una muestra aleatoria de 11 hombres de 17 años y 11 mujeres de 17 años, obtener la media muestral de las alturas de los dos grupos, y después realizar un test t de dos colas con $H_0: \mu_1 = \mu_2$ utilizando $\alpha = 0,05$. Podríamos esperar que 99 de los 100 investigadores encontraron evidencia estadísticamente significativa de que las alturas medias de los hombres y las mujeres de 17 años de edad son diferentes (es decir, evidencia significativa a favor de H_A). Podríamos esperar que uno de los 100 investigadores no encontrara suficiente evidencia sobre la diferencia, con el nivel de significación de 0,05. (Por tanto, un investigador cometería un error de Tipo II). ■

Como hemos visto, para elegir el tamaño de la muestra, el investigador necesita especificar no solo el tamaño del efecto que desea detectar, sino también con cuánta certeza desea detectarlo. Es decir, es necesario especificar cuánta potencia se desea. Como la potencia mide la protección contra el error de Tipo II, la elección del nivel de potencia deseado depende de las consecuencias que resultarían de un error de Tipo II. Si las consecuencias de un error de Tipo II fueran muy desafortunadas (por ejemplo, si un tratamiento del cáncer prometedor pero con riesgos se está probando en humanos y un resultado negativo desacreditaría el tratamiento, de forma que nunca se probaría de nuevo), entonces el

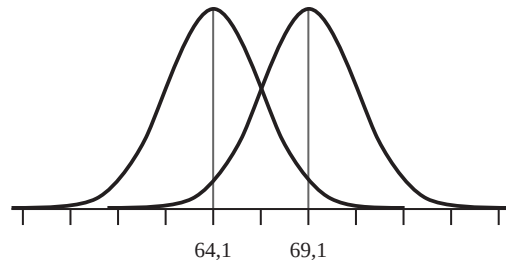


Figura 7.7.2 Distribución de alturas del Ejemplo 7.7.2

investigador debería especificar una potencia alta, por ejemplo 0,95 o 0,99. Pero, por supuesto, una potencia alta es costosa en términos de n . En muchas investigaciones, un error de Tipo II no es un desastre, y una potencia más baja como 0,80 se considera adecuada.

El ejemplo siguiente ilustra el uso típico de la Tabla 5 para planificar un experimento.

Ejemplo 7.7.3 Pérdida de peso posparto

Un grupo de científicos deseaban investigar si un programa de intervención basado en internet ayudaría o no a las mujeres a perder peso después de dar a luz. Un grupo de mujeres en la etapa de postparto se apuntaron a un programa basado en internet que proporcionaba orientación sobre alimentación y ejercicio semanal adecuada al tiempo transcurrido tras el parto, realizaba un seguimiento de su progreso en pérdida de peso y establecía un foro en línea sobre nutrición y comentarios de ejercicios con otras madres recientes. Otro grupo de mujeres en la etapa de postparto (el «grupo de control») recibió por parte de sus doctores instrucciones por escrito en la forma tradicional sobre la dieta y los ejercicios. La variable de respuesta del estudio fue la cantidad de pérdida de peso 12 meses después del parto en kg. Los estudios previos habían mostrado que 12 meses después del parto, la pérdida media de peso es de aproximadamente de 3,6 kg con una desviación típica de 4,0 kg. (Nota: una pérdida de peso negativa es una ganancia de peso). El equipo de investigación deseaba demostrar que había como mínimo un 50 % de mejora en la pérdida de peso en el grupo de internet. Es decir, desearía demostrar que las mujeres que siguieran el programa basado en internet perdían al menos 1,8 kg (el 50 % de 3,6 kg) más peso que las mujeres del grupo de control. Planearon realizar un test t de una cola con un nivel de significación del 5 %. El equipo tenía que decidir cuántas mujeres (n) asignar a cada grupo.

El tamaño del efecto que el equipo deseaba considerar es

$$\frac{|\mu_1 - \mu_2|}{\sigma} = \frac{1,8}{4,0} = 0,45$$

Para este tamaño del efecto, y para grupo una potencia de 0,80 con un test t de una cola y un nivel de significación del 5 %, la Tabla 5 da un resultado de $n = 62$, lo que significa que se necesitan 62 mujeres en cada grupo.

En este punto, el equipo de investigación tenía que considerar cuestiones tales como (1) ¿Es posible apuntar a 124 mujeres en la etapa de postparto (62 en cada grupo) para el estudio? Si no, entonces (2) ¿Se podría considerar quizá redefinir el tamaño de la diferencia entre los grupos que se considera importante, para reducir el valor de n requerido? Con cuestiones como esa, y un uso repetido de la Tabla 5, podrían finalmente decidir un valor firme de n , o en su caso decidir abandonar el proyecto porque un estudio adecuado resultaría demasiado costoso.

Normalmente la historia termina aquí, pero puede haber otra nueva vuelta de tuerca en la planificación del estudio. El equipo de investigación sabía de su experiencia que aproximadamente el 20 % de las mujeres que participan en este tipo de estudios terminan dejándolo, por una razón u otra, antes de finalizar el estudio. (No hay fórmula ni tabla que nos diga cuántos sujetos abandonarán en un estudio como este. Aquí la única guía es la experiencia). En este caso, el equipo de investigación planeó admitir a 150 mujeres (un poco más que el 20 % extra, 13 mujeres en cada grupo), para permitir algunos abandonos y terminar todavía con suficientes datos para tener la potencia que deseaban⁵¹.

Ejercicios 7.7.1-7.7.11

7.7.1 Una medida de la calidad de la carne de cerdo es el espesor de grasa dorsal. Suponga que dos investigadores, Jones y Smith, planean medir el espesor de grasa dorsal en dos grupos de cerdos que han seguido dietas diferentes. Han decidido utilizar el mismo número (n) de cerdos en cada grupo y comparar los espesores medios de grasa dorsal utilizando un test t de dos colas con un nivel de significación del 5 %. Los datos preliminares indican que la DT del espesor de grasa dorsal es aproximadamente 0,3 cm.

Cuando los investigadores consultaron a un estadístico para que les ayudara a elegir n , naturalmente les preguntó cuánta diferencia deseaban detectar, Jones respondió: «Si la verdadera diferencia es de 1/4 cm o más, deseo estar razonablemente seguro de rechazar H_0 ». Smith respondió: «Si la verdadera diferencia es de 1/2 cm o más, quiero estar muy seguro de rechazar H_0 ».

Si el estadístico interpreta «razonablemente seguro» como una potencia del 80 % y «muy seguro» como una potencia de 95 %, indique cuál es el valor de n que recomendaría:

- (a) Para satisfacer el requisito de Jones.
- (b) Para satisfacer el requisito de Smith.

7.7.2 Con referencia a los datos de NE del cerebro del Ejemplo 7.2.1, suponga que está planificando un experimento similar: estudiará el efecto del LSD (en vez del tolueno) sobre la NE en el cerebro. Va a utilizar un test t de dos colas con $\alpha = 0,05$. Suponga que ha decidido que un 10 % de efecto (aumento o disminución en el valor medio de NE) del LSD sería importante, y también que desea tener una buena potencia (80 %) para detectar una diferencia de esta magnitud.

- (a) Utilizando los datos del Ejemplo 7.2.1 como «estudio piloto», determine cuántas ratas debería haber en cada grupo. (El NE medio en el grupo de control del Ejemplo 7.2.1 es 444,2 ng/g y la DT es 69,6 ng/g).
- (b) Si estuviera planificando utilizar un test t de una cola, ¿cuál sería el número de ratas requerido?

7.7.3 Suponga que está planificando un experimento en invernadero sobre el crecimiento de plantas de pimiento. Suponga que cultiva n macetas en semillero con una tierra estándar y otras n macetas con una tierra tratada especialmente. Transcurridos 21 días, se mide Y = longitud total del tallo (cm) de cada planta. Si el efecto del tratamiento de la tierra es aumentar la longitud media del tallo de la población en 2 cm, se desea tener un 90 % de probabilidad de rechazar H_0 con un test t de una cola. Los datos de un estudio piloto (como los datos del Ejercicio 2.6.2) con 15 plantas cultivadas en tierra estándar son $\bar{y} = 12,5$ cm y $s = 0,8$ cm.

- (a) Suponga que planifica realizar el contraste con $\alpha = 0,05$. Utilice la información del estudio piloto para determinar qué valor de n habría que utilizar.
- (b) ¿Qué condiciones son necesarias para la validez del cálculo del apartado (a)? ¿Cuáles de ellas se pueden comprobar (aproximadamente) utilizando los datos del estudio piloto?

- (c) Suponga que decide adoptar una postura más conservadora y utilizar $\alpha = 0,01$. ¿Qué valor de n se debería utilizar?

7.7.4 Las medidas de la presión sanguínea diastólica en hombres americanos con edades comprendidas entre 18 y 44 años siguen aproximadamente una curva normal con $\mu = 81$ mmHg y $\sigma = 11$ mmHg. La distribución de las mujeres con edades entre 18 y 44 años es también aproximadamente normal con la misma DT pero con una media menor: $\mu = 75$ mmHg⁵². Suponga que se desea medir la presión sanguínea diastólica de n hombres seleccionados aleatoriamente y de n mujeres seleccionadas aleatoriamente con edades entre 18 y 44 años. Sea E el suceso de que un test t encontrará que la diferencia entre los hombres y las mujeres es estadísticamente significativa. Indique cuál debe ser el valor de n para que $\Pr\{E\} = 0,9$:

- (a) Si se utiliza un test t de dos colas con $\alpha = 0,05$.
- (b) Se utiliza un test t de dos colas con $\alpha = 0,01$.
- (c) Si se utiliza un test t de una cola (en la dirección correcta) con $\alpha = 0,05$.

7.7.5 Suponga que está planificando un experimento para probar el efecto de un cierto medicamento en el hábito de beber de las ratas. Utilizará un test t de dos colas para comparar un grupo de ratas que ha recibido el tratamiento con un grupo de control. La variable observada será Y = consumo de agua en una hora después de una privación de 23 horas. Suponga que ha decidido que, si el efecto de la medicina es desplazar el consumo medio de la población en 2 ml o más, desea tener un 80 % de probabilidad de obtener evidencia significativa a favor de H_A con un nivel de significación del 5 %.

- (a) Los datos preliminares indican que la DT de Y bajo las condiciones de control es aproximadamente de 2,5 ml. Utilizando esta información con una estimación preliminar de σ , determine cuántas ratas debería tener en cada grupo.
- (b) Suponga que, debido a que el cálculo del apartado (a) indica un número grande de ratas, considera modificar el experimento para reducir σ . Encuentra que, cambiando a un mejor suministrador de ratas y mejorando los procedimientos de laboratorio, podría reducir la DT a la mitad. Sin embargo, el coste de cada observación sería el doble. ¿Serían las nuevas medidas más efectivas en coste? Es decir, ¿serían el experimento modificado menos costoso?

7.7.6 Los datos de un amplio estudio indican que la concentración en suero de lactato deshidrogenasa (LD) es mayor en los hombres que las mujeres. (Los datos se resumen en el Ejemplo 7.6.1). Suponga que el doctor Sánchez se propone realizar su propio estudio para corroborar esta conclusión. Sin embargo, debido a la limitación de recursos, Sánchez solo puede conseguir 35 hombres y 35 mujeres para su estudio.

Suponiendo que la verdadera diferencia en las medias poblacionales es de 4 U/l y que la DT de cada población es de 10 U/l, ¿cuál es la probabilidad de que Sánchez tenga éxito? Concretamente, calcule la probabilidad de que Sánchez rechace H_0 utilizando un test t de una cola con un nivel de significación del 5 %.

7.7.7 En relación con el estudio sobre analgésicos del Ejercicio 7.5.10, recuerde que dicho estudio incluía 25 observaciones en cada grupo de tratamiento y que el tamaño del efecto era aproximadamente 0,5. Si este es el verdadero tamaño del efecto poblacional, ¿cuál es la probabilidad (aproximada) de obtener una diferencia significativa entre la efectividad media de los dos medicamentos en un experimento de este tamaño (es decir, cada muestra de tamaño 25)?

7.7.8 En relación con el estudio sobre analgésicos del Ejercicio 7.5.10, la evidencia en dicho estudio a favor del medicamento fue marginalmente significativa ($0,025 < P < 0,05$). Suponga que la doctora Williams está planificando un nuevo estudio sobre el mismo medicamento para intentar replicar el resultado original, es decir, demostrar que el medicamento es efectivo. La doctora considerará que este estudio tiene éxito si rechaza H_0 mediante un test t de una cola y $\alpha = 0,05$. En el estudio original, la diferencia entre las medias de los tratamientos era aproximadamente la mitad de media desviación típica [$(32 - 25)/13 \approx 0,5$]. Tomando esto como un valor provisional para el tamaño del efecto, determine cuántos pacientes debería tener Williams en cada grupo para que su probabilidad de éxito fuera

(a) 80 %

(b) 90 %

(Nota: Este problema ilustra el hecho de que se pueden necesitar tamaños de muestra sorprendentemente grandes para realizar un estudio de replicación que tenga éxito, especialmente si los resultados originales fueron solo marginalmente significativos).

7.7.9 Considere la comparación de dos distribuciones normales en las que el tamaño del efecto de la diferencia es

(a) 3

(b) 1

En cada caso, dibuje aproximadamente el aspecto del solapamiento de las distribuciones. (Véase la Figura 7.2.1).

7.7.10 Un científico de animales está planificando un experimento para evaluar un nuevo suplemento de dieta para el ganado. Un grupo de ganado recibirá una dieta estándar y un segundo grupo recibirá la dieta estándar más un suplemento. El investigador desea tener una potencia del 90 % para detectar un incremento en la ganancia media de peso de 20 kg, utilizando un test t de una cola con $\alpha = 0,05$. Basándose en su experiencia previa espera que la DT sea de 17 kg. ¿Cuántas cabezas de ganado necesita en cada grupo?

7.7.11 Una investigadora planea realizar un estudio que analizará con un test t de dos colas y el nivel de significación del 5 %. Puede permitirse reunir 20 observaciones en cada uno de los dos grupos de su estudio. ¿Cuál es el mínimo tamaño del efecto para el que tendrá al menos una potencia del 95 %?

7.8 t de Student: condiciones y resumen

En las secciones anteriores hemos presentado la comparación de dos medias utilizando métodos clásicos basados en la distribución t de Student. En esta sección presentaremos las condiciones en las que se basan estos métodos. Además, resumiremos los métodos para tener una referencia conveniente.

Condiciones

Los procedimientos del test t y el intervalo de confianza que hemos descrito son apropiados si se cumplen las siguientes condiciones*:

1. Condiciones sobre el diseño del estudio

- (a) Debe ser razonable considerar a los datos como muestras aleatorias de sus respectivas poblaciones. Las poblaciones deben ser grandes en relación con sus tamaños muestrales. Las observaciones dentro de cada muestra deben ser independientes.
- (b) Las dos muestras deben ser independientes entre sí.

2. Condiciones sobre la forma de las distribuciones poblacionales

Las distribuciones muestrales de $\bar{Y}_1$ e $\bar{Y}_2$ deben ser (aproximadamente) normales. Esto puede ser debido a la propia normalidad de las poblaciones o debido al Teorema Central del Límite (véase la Sección 6.5) si las poblaciones no son normales, pero los tamaños de la muestra son grandes, donde «grande» depende del grado de no normalidad de las poblaciones. En muchas situaciones prácticas, tamaños de muestra moderados (por ejemplo,

* Muchos autores utilizan la palabra «supuestos» donde nosotros utilizamos la palabra «condiciones».

$n_1 = 20, n_2 = 20$) son lo suficientemente «grandes». Sin embargo, siempre necesitamos tener en cuenta que uno o dos *outliers* pueden tener un efecto importante en los resultados de cualquier procedimiento estadístico, incluyendo el test *t*.

Verificación de condiciones

La comprobación de las condiciones anteriores debe ser una parte de cualquier análisis de datos.

La comprobación de la condición 1(a) sería con respecto al intervalo de confianza (Sección 6.5), el investigador buscaría sesgos en el diseño del experimento y verificaría que no hay estructuras jerárquicas dentro de cada muestra.

La condición 1(b) significa que no hay emparejamiento ni dependencia entre las dos muestras. El significado completo de esta condición resultará claro en los Capítulos 8 y 9.

Algunas veces se sabe de estudios previos si las poblaciones se pueden considerar aproximadamente normales. En ausencia de esta información, el requisito de normalidad se puede comprobar realizando a cada muestra por separado histogramas, gráficas de probabilidad normal o la prueba de normalidad de Shapiro-Wilk. Afortunadamente, el test *t* es bastante robusto frente desviaciones de la normalidad⁵³. Habitualmente, solo habrá que tener en cuenta desviaciones importantes de la normalidad debidas a *outliers*, o colas largas y desordenadas. El sesgo moderado tiene muy poco efecto en el test *t*, incluso para muestras pequeñas.

Consecuencias de un uso inapropiado de la *t* de Student

Esta presentación del test *t* y del intervalo de confianza (en las Secciones 7.3-7.8) está basada en las condiciones (1) y (2). La violación de esas condiciones puede hacer que los métodos resulten inapropiados.

Si las condiciones no se satisfacen, entonces el test *t* puede resultar inapropiado en dos formas posibles:

1. Puede ser inválido en el sentido de que el riesgo actual de cometer un error de Tipo I sea mayor que el nivel de significación nominal α . (Expresándolo de otra manera, el *P* valor del procedimiento del test *t* puede ser inapropiadamente pequeño).
2. El test *t* puede ser válido, pero con una potencia menor que un test más apropiado.

Si el diseño incluye estructuras jerárquicas que se han ignorado en el análisis, el test *t* puede ser seriamente inválido. Si las muestras no son independientes entre sí, la consecuencia habitual es una pérdida de potencia.

Una forma bastante común de apartarse de la condición de normalidad sucede cuando una o ambas poblaciones tienen colas largas y desordenadas. El efecto de esta forma de no normalidad es aumentar el ET y, por tanto, restar potencia al test *t*.

El uso inadecuado de los intervalos de confianza es análogo al de los tests *t*. Si se violan las condiciones, entonces el intervalo de confianza puede no ser válido, es decir, demasiado estrecho para el nivel de confianza especificado, o puede ser válido pero más ancho de lo necesario.

Otros planteamientos

Como los métodos basados en la distribución *t* de Student pueden no ser siempre los más apropiados, los estadísticos han desarrollado otros métodos que tienen propósitos similares. Uno de ellos es el test de Wilcoxon-Mann-Whitney, que presentaremos en la Sección 7.10. Otra posible solución a la dificultad es transformar los datos, por ejemplo, analizar $\log(Y)$ o $\ln(Y)$ en vez de *Y*.

Ejemplo 7.8.1 Inflamación de tejidos

Unos investigadores tomaron muestras de piel de 10 pacientes que tenían implantes de pecho y de un grupo de control de 6 pacientes. Tras cultivar las muestras de tejido durante 24 horas, apuntaron los niveles de interleucina-6 (en pg/ml/10 g de tejido), una medida de la inflamación de los tejidos. La Tabla 7.8.1 muestra los datos⁵⁴. La Figura 7.8.1(a) muestra el diagrama de puntos en paralelo de esos datos. Las gráficas de probabilidad normal que se muestran en la Figura 7.8.2(a) indican que las distribuciones están severamente sesgadas, por lo que es necesaria una transformación antes de poder utilizar el procedimiento de la *t* de Student. Tomando el logaritmo en base 10 de cada observación se obtienen los valores que se muestran en las columnas de la parte derecha de la Tabla 7.8.1 y en la

Figura 7.8.1(b). La gráfica de probabilidad normal de la Figura 7.8.2(b) muestra que, con los datos transformados a escala logarítmica, se cumple la condición de normalidad. Por tanto, realizaremos un análisis de los datos en escala logarítmica. Es decir, contrastaremos

$$H_0: \mu_1 = \mu_2$$

frente a

$$H_A: \mu_1 \neq \mu_2$$

Tabla 7.8.1 Niveles de interleucina-6 en pacientes con implantes de pecho y en pacientes de control				
	Datos originales		Escala logarítmica	
	Pacientes con implantes de pecho	Pacientes de control	Pacientes con implantes de pecho	Pacientes de control
	231	35.324	2,364	4,548
	308.287	12.457	5,489	4,095
	33.291	8.276	4,522	3,918
	124.550	44	5,095	1,643
	17.075	278	4,232	2,444
	22.955	840	4,361	2,924
	95.102		4,978	
	5.649		3,752	
	840.585		5,925	
	58.924		4,770	
$\bar{y}$	150.665	9.537	4,549	3,262
s	259.189	13.613	0,992	1,111

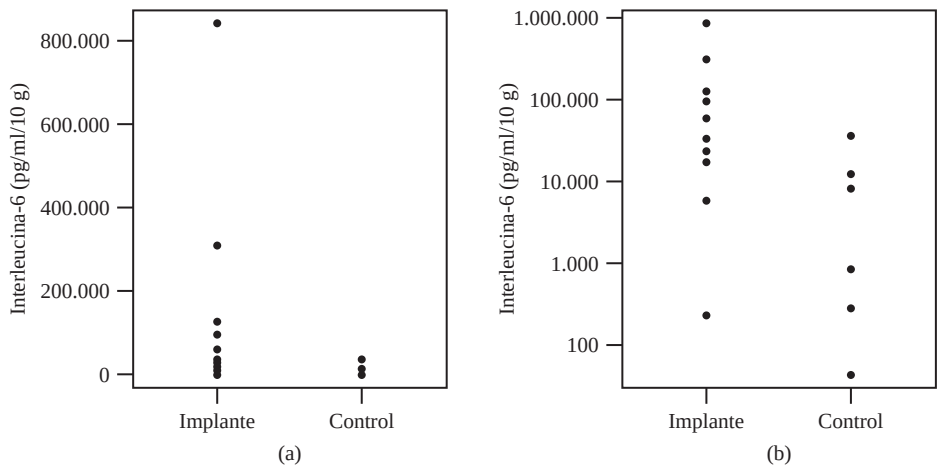


Figura 7.8.1 Diagramas de puntos de los datos de inflamación de tejidos del Ejemplo 7.8.1 (a) en la escala original; (b) en escala logarítmica

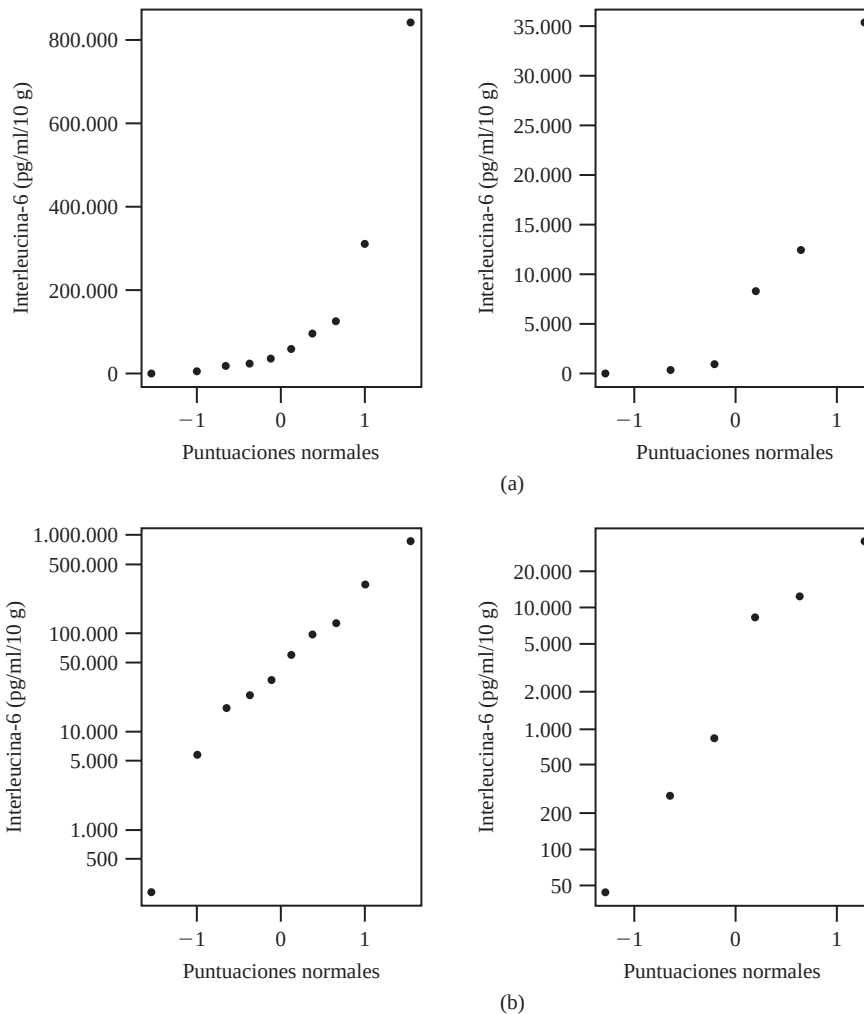


Figura 7.8.2 Gráficas de probabilidad normal de los datos de inflamación de tejidos del Ejemplo 7.8.1 (a) en la escala original; (b) en escala logarítmica

siendo μ_1 la media poblacional del logaritmo del nivel de interleucina-6 de las pacientes con implantes de pecho y μ_2 la media poblacional del logaritmo del nivel de interleucina-6 para los pacientes de control. Supongamos que elegimos $\alpha = 0,10$. El estadístico de contraste es

$$t_s = \frac{(4,549 - 3,262)}{0,553} = 2,33$$

La fórmula (6.7.1) da como resultado $gl = 9,7$. El P valor de este contraste es 0,045. Por tanto, tenemos evidencia, con un nivel de significación de 0,10 (y también con un nivel de 0,05), de que la media del logaritmo del nivel de interleucina-6 es mayor en la población con implantes de pecho que en la población de control. ■

Resumen del procedimiento del test *t*

Con objeto de tener una referencia conveniente, resumimos el procedimiento del test basado en la *t* de Student para contrarrestar la igualdad de medias de muestras independientes.

Test t

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2 \text{ (no direccional)}$$

$$H_A: \mu_1 < \mu_2 \text{ (direccional)}$$

$$H_A: \mu_1 > \mu_2 \text{ (direccional)}$$

$$\text{Estadístico de contraste: } t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\text{ET}_{(\bar{y}_1 - \bar{y}_2)}}$$

P valor = área de la cola bajo la curva t de Student con

$$gl = \frac{(ET_1^2 + ET_2^2)^2}{ET_1^4/(n_1 - 1) + ET_2^4/(n_2 - 1)}$$

H_A no direccional: P valor = área en las dos colas más allá de t_s y $-t_s$

H_A direccional: Paso 1. Comprobar la direccionalidad.

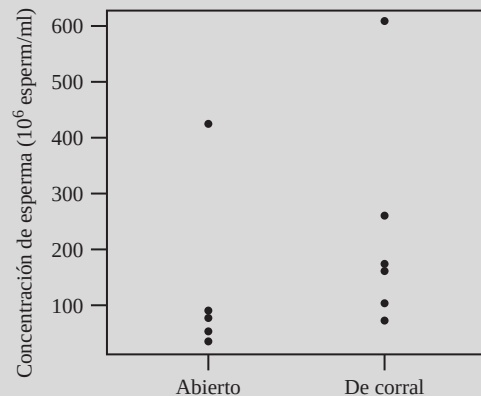
Paso 2. P valor = área en una sola cola más allá de t_s

Decisión: evidencia significativa a favor de H_A si P valor $\leq \alpha$.

Ejercicios 7.8.1-7.8.2

7.8.1 Para determinar si el entorno puede afectar a la calidad y producción de esperma en el ganado, un investigador asignó aleatoriamente 13 toros a dos entornos. 6 fueron criados en un entorno abierto y los otros 7 fueron criados en un entorno de corral más pequeño. El diagrama siguiente muestra las concentraciones de esperma (millones de espermatozoides/ml) de muestras de semen de los 13 toros⁵⁵.

- Utilizando la gráfica anterior para justificar su respuesta, ¿sería apropiado utilizar el método de la t de Student para comparar las concentraciones medias de esperma bajo estas dos condiciones experimentales?
- ¿Cómo sería su respuesta al apartado (a) si los datos fueran de 60 y 70 especímenes en vez de 6 y 7?
- La prueba de normalidad de Shapiro-Wilk da un P valor de 0,0012 y 0,0139 para los datos de abierto y corral, respectivamente. ¿Cómo dan soporte o refutan estos resultados su respuesta del apartado (a)?
- ¿Cómo podría ayudar una transformación a analizar estos datos?



7.8.2 En referencia a los datos sobre serotonina del Ejercicio 7.27, ¿sobre qué base podría ponerse una objeción al uso del test t con estos datos? (*Sugerencia:* calcule la DT de cada muestra y compárela con la media muestral).

7.9 Más sobre los principios del contraste de hipótesis

Nuestro estudio del test t ha ilustrado alguno de los principios generales de los contrastes de hipótesis en estadística. En el resto de este libro presentaremos algunos otros tipos de contrastes además del test t .

Una visión general de los contrastes de hipótesis

En un contraste de hipótesis estadístico intervienen una hipótesis nula H_0 , una hipótesis alternativa o hipótesis de investigación, H_A , y un estadístico de contraste que mide la desviación o discrepancia de los datos con respecto a H_0 . La distribución muestral del estadístico de contraste, bajo el supuesto de que H_0 es cierta, se denomina **distribución nula** del

estadístico de contraste. (Si estamos realizando un test de aleatorización como en la Sección 7.1, entonces la distribución nula es la distribución de todas las posibles diferencias en las medias muestrales debidas a la asignación aleatoria de observaciones a grupos, como se muestra en la Tabla 7.1.2; como otro ejemplo, si estamos realizando un test t , entonces la distribución nula del estadístico de contraste t_s es, bajo ciertas condiciones, una distribución t de Student). La distribución nula indica cuánto se puede esperar que el estadístico de contraste se desvíe con respecto a H_0 debido únicamente al azar.

Al contrastar una hipótesis, evaluamos la evidencia en contra de H_0 (y a favor de H_A) situando el estadístico de contraste dentro de la distribución nula. El P valor es una medida de dicha situación, que indica el grado de compatibilidad entre los datos y H_0 . La línea divisoria entre la compatibilidad y la incompatibilidad queda especificada por un nivel de significación α elegido arbitrariamente. La decisión sobre afirmar que existe evidencia significativa a favor de H_A se hace de acuerdo a la siguiente regla:

Rechazar H_0 si P valor $\leq \alpha$.

Cuando no se dispone de computador, no podemos calcular exactamente el P valor, pero podemos acotarlo utilizando una tabla de valores críticos. Si H_A es direccional, la acotación del P valor es un procedimiento de dos pasos.

Todo contraste de hipótesis nula H_0 tiene asociados riesgos de error de Tipo I (obtener evidencia significativa a favor de H_A cuando H_0 es cierta) y de Tipo II (no encontrar evidencia significativa a favor de H_A cuando H_A es cierta). El riesgo del error de Tipo I está siempre limitado por el nivel de significación elegido, α :

$\Pr\{\text{Rechazar } H_0\} \leq \alpha$ si H_0 es cierta

Por tanto, el procedimiento del contraste de hipótesis trata al error de Tipo I como aquel del que hay que protegerse en mayor medida. Por el contrario, la potencia de un contraste puede ser bastante baja y, de forma equivalente, el riesgo de un error de Tipo II puede ser bastante grande, si las muestras son pequeñas.

¿Cómo se eligen H_0 y H_A ?

Una dificultad común cuando se estudian los contrastes de hipótesis por primera vez es comprender cómo deben ser las hipótesis nula y alternativa. En general, la hipótesis nula representa el status quo (lo que se debería creer, por defecto, a menos que los datos demuestren otra cosa)*. En general, la hipótesis alternativa es la afirmación que el investigador intenta establecer. Por tanto, H_A se denomina también hipótesis de investigación. Por ejemplo, si estamos probando un nuevo medicamento frente a un medicamento estándar, la hipótesis de investigación es que el nuevo medicamento es mejor que el medicamento estándar, mientras que la hipótesis nula es que el nuevo medicamento no es diferente del estándar (en ausencia de evidencia, esperamos que los dos medicamentos sean igual de efectivos). La hipótesis nula típica, $H_0: \mu_1 = \mu_2$, establece que las dos medias poblacionales son iguales y que cualquier diferencia entre las medias muestrales es simplemente debida al error aleatorio en el proceso de muestreo. La hipótesis alternativa es que existe una diferencia entre los medicamentos, de forma que cualquier diferencia observada en las medias muestrales es debida a un efecto real, en vez de ser debida solo al error aleatorio. Concluimos que tenemos evidencia estadísticamente significativa a favor de la hipótesis de investigación si los datos muestran una diferencia en las medias muestrales más allá de la que puede ser atribuida razonablemente al azar.

He aquí otros ejemplos: si estamos comparando hombres y mujeres con respecto a algún atributo, la hipótesis nula habitual es que no hay diferencia, en promedio, entre los hombres y las mujeres. Si estamos estudiando una medida de biodiversidad en dos entornos, la hipótesis nula habitual es que las biodiversidades de los dos entornos son iguales, en promedio. Si estamos estudiando dos dietas, la hipótesis nula habitual es que las dos dietas producen la misma respuesta en promedio.

Otro vistazo al P valor

Para situar el P valor en un contexto general, consideremos algunas de sus interpretaciones verbales.

Revisaremos primero el test de aleatorización. Con una H_A no direccional, el P valor es la proporción de todas las aleatorizaciones que producen una diferencia de las medias muestrales que es tan grande como, o mayor que, la diferencia que se observó en el estudio real. Por tanto, podemos definir el P valor como sigue:

* Esta regla general no es siempre cierta, se ofrece solo como orientación.

El P valor de los datos es la probabilidad (suponiendo que H_0 es cierta) de obtener un resultado tan extremo como, o más extremo que, el resultado que fue realmente observado.

Expresándolo de otra forma:

El P valor es la probabilidad de que, si H_0 fuera cierta, se obtuviera un resultado que se desviará con respecto a H_0 tanto como (o más que) lo hacen los datos reales.

Consideremos ahora el test t . Dada una H_A no direccional, hemos definido el P valor como el área en las dos colas bajo la curva t de Student más allá del valor observado de t_s .

Realmente, estas descripciones del P valor son un tanto limitadas. El P valor depende realmente de la naturaleza de la hipótesis alternativa. Cuando estamos realizando un test t contra una alternativa *direccional*, el P valor de los datos es (si la desviación observada está en la dirección de H_A) solo el área de *una sola cola* más allá del valor observado de t_s . La definición más general de P valor es la siguiente:

El P valor de los datos es la probabilidad (asumiendo que H_0 es cierta) de obtener un resultado que se desvía tanto como, o que se desvía más que, el resultado observado realmente. La desviación se mide como discrepancia con respecto a H_0 en la dirección de H_A .

El P valor mide la facilidad con que la desviación observada se puede explicar como una variación aleatoria en vez de por la explicación alternativa proporcionada por H_A . Por ejemplo, si el test t da un P valor de $P = 0,036$ para nuestros datos, entonces podemos decir que si H_0 fuera cierta esperaríamos que los datos se desviarán con respecto a H_0 tanto como lo hicieron nuestros datos solo un 3,6 % de las veces (en el metaestudio).

Otra definición de P valor sobre la que merece la pena pensar es la siguiente:

El P valor de los datos es el valor de α para el que se rechazaría H_0 en el límite, utilizando esos datos.

Para interpretar esta definición, imaginemos que un informe de investigación que incluye un P valor es leído por varios científicos interesados. Los científicos que son muy escépticos sobre H_A requerirían una evidencia muy fuerte antes de quedar convencidos y, por tanto, utilizarían un umbral de decisión muy conservativo, como por ejemplo $\alpha = 0,001$. Los científicos que están más favorablemente dispuestos hacia H_A requerirían solo una evidencia débil y, por tanto, utilizarían un valor más amplio como, por ejemplo, $\alpha = 0,10$. El P valor de los datos determina el punto, dentro de este espectro de opinión, que separa aquellos que encuentran que los datos son convincentes a favor de H_A y aquellos que no. Por supuesto, si el P valor es grande, por ejemplo $P = 0,40$, entonces presumiblemente ninguna persona razonable rechazaría H_0 y estaría convencido de H_A .

Como muestra de la explicación anterior, el P valor no describe todas las facetas de los datos, sino que se relaciona solo con un contraste de una hipótesis nula particular contra una alternativa particular. De hecho, veremos que el P valor de los datos también depende de qué contraste estadístico se utiliza para probar una hipótesis nula dada. Por esta razón, cuando en un informe científico se presentan los resultados de un contraste estadístico, es mejor incluir el P valor (exacto, si es posible), el nombre del contraste estadístico y si la hipótesis alternativa era direccional o no direccional.

Repetimos aquí, dado que se aplica a cualquier contraste estadístico, el principio expuesto en la Sección 7.6: el P valor es una medida de la fuerza de la evidencia contra H_0 , pero el P valor no refleja la *magnitud* de la discrepancia entre los datos y H_0 . Los datos se pueden desviar con respecto a H_0 solo ligeramente y, si las muestras son grandes, el P valor todavía puede ser bastante pequeño. Por el mismo motivo, los datos que se desvían sustancialmente con respecto a H_0 nunca pueden producir un P valor grande. El P valor por sí mismo no indica si un descubrimiento científico es importante.

Interpretación de las probabilidades de error

Un error común es interpretar el P valor como la probabilidad de que la hipótesis nula sea cierta. Un error relacionado es creer que, si obtenemos evidencia significativa a favor de H_A (por ejemplo) con un nivel de significación del 5 %, entonces la probabilidad de que H_0 sea cierta es del 5 %. Estas interpretaciones no son correctas*. Este punto se puede ilustrar con una analogía en el diagnóstico médico.

* De hecho, la probabilidad de que H_0 sea cierta no se puede calcular en absoluto con el planteamiento «frecuentista» estándar del contraste de hipótesis. $\Pr\{H_0 \text{ sea cierta}\}$ se puede calcular utilizando lo que se denominan métodos bayesianos, que están fuera del alcance de este libro.

Al aplicar una prueba diagnóstica de una enfermedad, la hipótesis nula es que la persona está sana (esto es lo que creemos a menos que la prueba diagnóstica indique lo contrario). Son posibles dos tipos de error: un individuo sano puede ser diagnosticado enfermo (falso positivo) y un individuo enfermo puede ser diagnosticado sano (falso negativo). El realizar la prueba diagnóstica sobre individuos que se sabe que son sanos o que están enfermos puede facilitarnos una estimación de la proporción de esos grupos cuando se diagnostican incorrectamente. Pero esta información por sí misma no nos dirá qué proporciones de todos los diagnósticos positivos son diagnósticos falsos. Estas ideas se ilustran numéricamente con el siguiente ejemplo.

Ejemplo 7.9.1 Pruebas médicas

Supongamos que se realiza una prueba médica para detectar una enfermedad. Supongamos además que el 1 % de la población tiene la enfermedad en cuestión. Si la prueba indica que la enfermedad está presente, rechazaremos la hipótesis nula de que la persona está sana. Si H_0 es cierta, entonces esto es un error de Tipo I (un falso positivo) y si la prueba indica que la enfermedad no está presente, tenemos una falta de evidencia significativa a favor de H_A (enfermedad). Supongamos que el contraste tiene un 80 % de probabilidad de detectar la enfermedad si una persona la tiene (esto es análogo a decir que la potencia del contraste de hipótesis es del 80 %) y un 95 % de probabilidades de indicar correctamente que la enfermedad está ausente si la persona realmente no la tiene (esto es análogo a decir que hay un 5 % de probabilidad de error de Tipo I). La Figura 7.9.1 muestra un árbol de probabilidades de esta situación, en el que las líneas gruesas indican las dos formas en las que la prueba puede dar positiva (es decir, las dos formas en las que se puede rechazar H_0).

Supongamos ahora que se realiza la prueba a 100.000 personas y que 1.000 de ellas (1 %) tienen realmente la enfermedad. Entonces, podemos esperar resultados como los que se muestran en la Tabla 7.9.1, con 5.750 personas que dan un resultado positivo (que es lo mismo que encontrar evidencia significativa a favor de H_A 5.750 veces). De ellas, 4.950 son falsos positivos. Expresándolo de otra forma, la proporción de las veces que H_0 es cierta, sabiendo que hemos encontrado evidencia significativa a favor de H_0 , es $\frac{4.950}{5.750} \approx 0,86$, que es bastante diferente de 0,05. Esta proporción alarmantemente alta de falsos positivos es debido a la rareza de la enfermedad. (La proporción de veces que ha existido evidencia significativa a favor de H_A , sabiendo que H_0 es cierta, es $\frac{4.950}{99.000} \approx 0,056$, tal como se

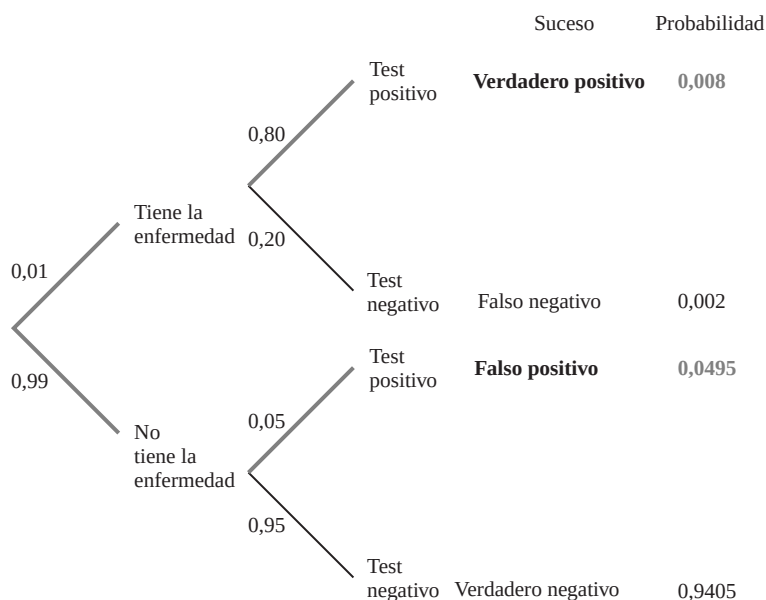


Figura 7.9.1 Árbol de probabilidades del ejemplo de la prueba médica

esperaba, pero se trata de una probabilidad condicional diferente. $\Pr\{A \text{ dado } B\} \neq \Pr\{B \text{ dado } A\}$: la probabilidad de lluvia, suponiendo que hay truenos y relámpagos, no es la misma que la probabilidad de que haya truenos y relámpagos, suponiendo que llueva). ■

Tabla 7.9.1 Resultados hipotéticos de la prueba médica de 100.000 personas				
		Situación verdadera		Total
		Sano (H_0 verdadero)	Enfermo (H_A verdadero)	
RESULTADO DE LA PRUEBA	Negativo (falta de evidencia significativa a favor de H_A)	94.050	200	94.250
	Positivo (evidencia significativa a favor de H_A)	4.950	800	5.750
Total		99.000	1.000	100.000

El riesgo de un error de Tipo I es una probabilidad que se calcula **bajo el supuesto de que H_0 es cierta**. De forma similar, el riesgo de un error de Tipo II se calcula suponiendo que H_A es cierta. Si tenemos un estudio bien diseñado, con tamaños de muestra adecuados, ambas probabilidades serán pequeñas. Tenemos entonces un buen procedimiento de contraste en el mismo sentido en el que la prueba médica es un buen procedimiento diagnóstico. Pero eso en sí mismo no garantiza que la mayor parte de las hipótesis nulas que rechazamos sean de verdad falsas, ni que la mayor parte de las que no rechazamos sean de hecho verdaderas. La validez o no validez de tales garantías dependerían de una cantidad desconocida y que no es posible conocer (concretamente, la proporción de hipótesis nulas verdaderas entre todas las hipótesis nulas que se contrastan, que es análogo a la incidencia de la enfermedad en el escenario de la prueba médica).

Perspectiva

Deberíamos mencionar que la filosofía del contraste de hipótesis estadístico que hemos explicado en este capítulo no es compartida por todos los estadísticos. La perspectiva presentada aquí, que se denomina **perspectiva frecuentista**, es ampliamente utilizada en investigación científica. Una perspectiva alternativa, la **perspectiva bayesiana**, incorpora no solo los datos observados en un estudio que se está realizando, sino también la información que el investigador tiene de estudios previos relacionados. En el pasado, muchas técnicas bayesianas no eran prácticas debido a la complejidad de las matemáticas que requieren. Sin embargo, la mayor potencia de cálculo y las mejoras en el *software* han hecho que los métodos bayesianos sean más populares en los últimos años.

Ejercicio 7.9.1

7.9.1 Suponga que se ha realizado un test t con $\alpha = 0,05$, y el P valor es 0,04. Indique si cada una de las afirmaciones siguientes es verdadera o falsa y explique por qué.

(a) Hay un 4 % de probabilidad de que sea cierta.

(b) Rechazamos H_0 con $\alpha = 0,05$.

(c) Deberíamos rechazar H_0 , y si repitiéramos el experimento, hay un 4 % de probabilidad de que rechazamos H_0 de nuevo.

(d) Si H_0 es cierta, la probabilidad de tener un estadístico de contraste que sea al menos tan extremo como el valor de t_s que se obtuvo realmente es del 4 %.

7.10 El test de Wilcoxon-Mann-Whitney

El **test de Wilcoxon-Mann-Whitney** se utiliza para comparar dos muestras independientes*. Es un competidor del test t , pero, a diferencia del test t , el test de Wilcoxon-Mann-Whitney es válido incluso si las distribuciones poblacionales

* El test presentado aquí fue desarrollado por Wilcoxon en un artículo de 1945. Mann y Whitney, en un artículo de 1947, elaboraron el test, que se puede realizar de dos formas matemáticamente equivalentes. Por tanto, algunos libros y programas estadísticos implementan el test de una forma diferente a la presentada aquí. Nótese también que algunos libros se refieren a este test como el test de Wilcoxon, algunos otros como el test de Mann-Whitney y algunos otros (incluyendo este texto) como el test de Wilcoxon-Mann-Whitney.

no son normales. Se dice por tanto que el test de Wilcoxon-Mann-Whitney es un test **independiente de la distribución**. Además, el test de Wilcoxon-Mann-Whitney no se enfoca en ningún parámetro particular como la media o la mediana. Por esta razón se dice que es un tipo de test **no paramétrico**.

Planteamiento de H_0 y H_A

Denotemos como Y_1 e Y_2 las observaciones de las dos muestras. Un planteamiento general de las hipótesis nula y alternativa de un test Wilcoxon-Mann-Whitney es

H_0 : Las distribuciones poblacionales de Y_1 e Y_2 son las mismas.

H_A : La distribución poblacional de Y_1 está desplazada con respecto a la distribución poblacional de Y_2 (es decir, Y_1 tiende a ser o más grande o más pequeña que Y_2).

En la práctica, es más natural plantear H_0 y H_A con las palabras adecuadas a la aplicación concreta, como se ilustra en el Ejemplo 7.10.1.

Ejemplo 7.10.1 Respiración del suelo

La respiración del suelo es una medida de la actividad microbiana en el mismo, que afecta al crecimiento de las plantas. En un estudio se tomaron muestras de suelo de dos localizaciones en un bosque: (1) en un claro del bosque (la localización «claro») y (2) en un área cercana con una alta densidad de árboles (la localización «densa»). Se midió la cantidad de dióxido de carbono liberada por cada muestra de suelo (en mol CO_2/g de suelo/hora). La Tabla 7.10.1 contiene los datos⁵⁶.

Una hipótesis nula apropiada se podría plantear como

H_0 : Las poblaciones de las que se extrajeron las dos muestras tienen la misma distribución de respiración del suelo.

Tabla 7.10.1 Datos de respiración del suelo (mol CO_2/g del suelo/hora) del Ejemplo 7.10.1									
Densa					Claro				
17	20	170	315		22	29	13	16	
22	190	64			15	18	14	6	

o, más informalmente, como

H_0 : Las áreas claro y densa no difieren con respecto a la respiración del suelo.

Una alternativa no direccional se podría plantear como

H_A : La distribución de las tasas de respiración del suelo tiende a ser mayor en una de las dos poblaciones.

O la hipótesis alternativa podría ser direccional, por ejemplo,

H_A : Las tasas de respiración del suelo tienden a ser mayores en el área «densa» que en el área «clara». ■

Aplicabilidad del test de Wilcoxon-Mann-Whitney

La Figura 7.10.1 muestra los diagramas de puntos de los datos de respiración del suelo del Ejemplo 7.10.1. La Figura 7.10.2 muestra las gráficas de probabilidad normal de estos datos. La distribución del área «densa» está sesgada hacia la derecha, mientras que la distribución del área «clara» está ligeramente sesgada hacia la izquierda. Si ambas distribuciones estuvieran sesgadas hacia la derecha, podríamos aplicar una transformación a los datos. Sin embargo, cualquier intento de transformar la distribución del área «densa», como tomar el logaritmo de los datos, hará que el sesgo de la distribución del área «clara» empeore. Por tanto, el test t no es aplicable aquí. El test de Wilcoxon-Mann-Whitney no requiere normalidad de las distribuciones.

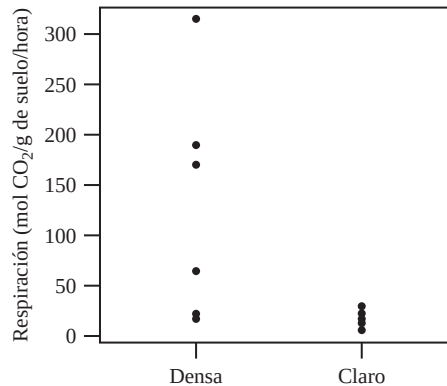


Figura 7.10.1 Diagramas de puntos de los datos de respiración del suelo del Ejemplo 7.10.1

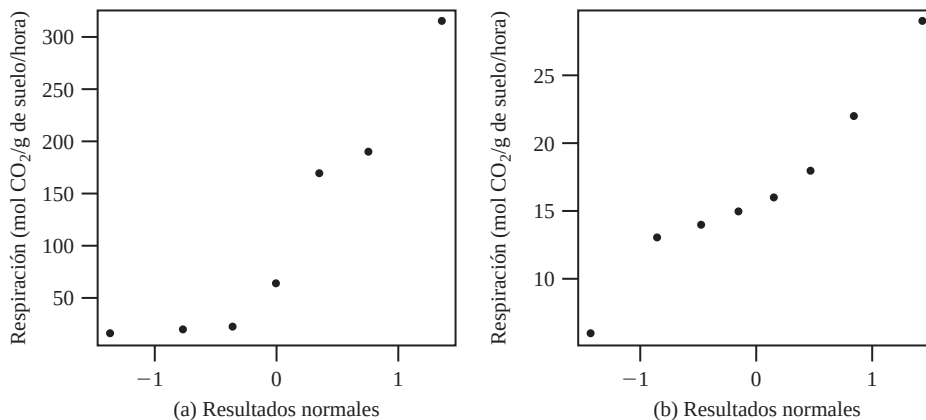


Figura 7.10.2 Gráficas de probabilidad normal de (a) los datos de densa y (b) los datos de claro del Ejemplo 7.10.1

Método

El estadístico de contraste de Wilcoxon-Mann-Whitney, que se denomina U_s , mide el grado de separación o desplazamiento entre las dos muestras. Un valor grande de U_s indica que las dos muestras están bien separadas, con relativamente poco solapamiento entre ellas. Los valores críticos del test de Wilcoxon-Mann-Whitney se muestran en la Tabla 6 al final del libro. El ejemplo siguiente ilustra el test de Wilcoxon-Mann-Whitney.

Ejemplo 7.10.2 Respiración del suelo

Realicemos un test de Wilcoxon-Mann-Whitney sobre los datos de diversidad del Ejemplo 7.10.1.

1. El valor de U_s depende de las posiciones relativas de los valores de Y_1 e Y_2 . El primer paso para determinar U_s es disponer las observaciones en orden creciente, como muestra la Tabla 7.10.2.
2. Seguidamente, determinaremos dos valores, K_1 y K_2 , como sigue:
 - (a) *El valor de K_1 .* Para cada observación de la muestra 1, contaremos el número de observaciones de la muestra 2 cuyo valor es menor (es decir, a la izquierda). Sumaremos $1/2$ para cada observación empatada. En los datos anteriores, hay cinco valores de Y_2 menores que el primer Y_1 , hay seis valores de Y_2 menores que el segundo Y_1 , hay seis valores menores que el tercer Y_1 y uno igual, por lo que el valor del recuento es $6 + 1/2$. Hasta el momento, los valores obtenidos son 5, 6, y 6,5. Continuando de forma similar,

Tabla 7.10.2 Cálculos de Wilcoxon-Mann-Whitney para el Ejemplo 7.10.2

Número de observaciones de claro cuyo valor es menor	Y_1 Datos de densa	Y_2 Datos de claro	Número de observaciones de densa cuyo valor es menor
5	17	6	0
6	20	13	0
6,5	22	14	0
8	64	15	0
8	170	16	0
8	190	18	1
8	315	22	2,5
		29	3
$K_1 = 49,5$		$K_2 = 6,5$	

obtendríamos los valores 8, 8, 8, y 8. En conjunto hay siete valores, uno para cada valor de Y_1 . La suma de los siete valores es $K_1 = 49,5$.

- (b) *El valor de K_2* . Para cada observación de la muestra 2, contaremos el número de observaciones de la muestra 1 cuyo valor es menor, sumando 1/2 para cada observación empatada. Esto da como resultado los valores de 0, 0, 0, 0, 0, 1, 2,5 y 3. La suma de sus valores es $K_2 = 6,5$.
- (c) *Comprobación*. Si el trabajo es correcto, la suma de K_1 y K_2 debe ser igual al producto de los tamaños de las muestras:

$$K_1 + K_2 = n_1 n_2$$

$$49,5 + 6,5 = 7 \times 8$$

3. El estadístico de contraste U_s es el mayor valor de K_1 y K_2 . En este ejemplo, $U_s = 49,5$.
4. Para determinar el P valor, consultamos la Tabla 6 con $n =$ el tamaño muestral más grande, y $n' =$ el tamaño muestral más pequeño. En el presente caso, $n = 8$ y $n' = 7$. Los valores de la Tabla 6 se reproducen en la Tabla 7.10.3.

Tabla 7.10.3 Valores de la Tabla 6 para $n = 8$, $n' = 7$

40	0,189	44	0,093	46	0,054	47	0,040	48	0,021	49	0,014	50	0,009

Contrastemos H_0 contra una alternativa no direccional con un nivel de significación $\alpha = 0,05$. En la Tabla 7.10.3, podemos ver que cuando $U_s = 49$, el P valor es 0,014 y cuando $U_s = 50$, el P valor es 0,009. Como $49 < U_s < 50$, el P valor está entre 0,009 y 0,014 y, por tanto, existe evidencia significativa a favor de H_A . Hay suficiente evidencia para concluir que las tasas de respiración del suelo son diferentes en las áreas «clara» y «densa». ■

Como ilustra el Ejemplo 7.10.2, la Tabla 6 se puede utilizar para acotar el P valor del test de Wilcoxon-Mann-Whitney del mismo modo que se utilizaba la Tabla 4 en el test t . Si el valor observado de U_s no aparece en la tabla, simplemente se localizan los valores que acotan al U_s observado. Después se acota el P valor mirando las correspondientes cabeceras de columna.

Direccionalidad En el test t , se determina la direccionalidad de los datos viendo si $\bar{Y}_1 > \bar{Y}_2$ o $\bar{Y}_1 < \bar{Y}_2$. De forma similar, se puede comprobar la direccionalidad en el test de Wilcoxon-Mann-Whitney comparando K_1 y K_2 : $K_1 > K_2$ indica que los valores de $\bar{Y}_1$ tienden a ser mayores que los de $\bar{Y}_2$, mientras que $K_1 < K_2$ indica la tendencia opuesta. A menudo, sin embargo, esta comparación formal no es necesaria y basta con echar un vistazo a la gráfica de los datos.

Alternativa direccional Si la hipótesis alternativa H_A es direccional en vez de no direccional, hay que modificar el procedimiento de Wilcoxon-Mann-Whitney. Como en el test t , el procedimiento modificado tiene dos pasos y el segundo paso requiere dividir por dos el P valor no direccional para obtener el P valor direccional.

Paso 1 Comprobar la direccionalidad, es decir, ver si los datos se desvían con respecto a H_0 en la dirección especificada por H_A .

- (a) Si no es así, el P valor es mayor que 0,50.
- (b) Si es así, continuar con el paso 2.

Paso 2 El P valor de los datos es la mitad del que sería si H_A fuera no direccional.

Para tomar una decisión con un nivel de significación especificado α , afirmamos que existe evidencia significativa a favor de H_A si P valor $\leq \alpha$.

El ejemplo siguiente ilustra el procedimiento en dos pasos.

Ejemplo 7.10.3 H_A direccional

Supongamos que $n = 8$, $n' = 7$, y H_A es direccional. Supongamos además que los datos se desvían con respecto a H_0 en la dirección especificada por H_A . Los valores que se muestran en la Tabla 7.10.3 se pueden utilizar para obtener el P valor como sigue:

- Si $U_s = 40$, entonces P valor = $0,189/2 = 0,0945$.
- Si $U_s = 46$, entonces P valor = $0,054/2 = 0,027$.
- Si $U_s = 49,5$, entonces $0,009/2 < P$ valor $< 0,014/2$ por lo que $0,0045 < P$ valor $< 0,007$.
- Si $U_s = 50$ (o máximo), entonces P valor $< 0,009/2 = 0,0045$. ■

Fundamento

Veamos por qué el procedimiento de contraste de Wilcoxon-Mann-Whitney tiene sentido. Para usar un caso específico, supongamos que los tamaños muestrales son $n_1 = 5$ y $n_2 = 4$, de forma que hay $5 \times 4 = 20$ comparaciones que se pueden realizar entre un valor de los datos de la primera muestra con otro valor de los datos de la segunda muestra. Por tanto, independientemente del aspecto de los datos, debemos tener

$$K_1 + K_2 = 5 \times 4 = 20$$

Las magnitudes relativas de K_1 y K_2 indican la cantidad del solapamiento de los valores de $\bar{Y}_1$ e $\bar{Y}_2$. La Figura 7.10.3 muestra cómo funciona. Para los datos de la Figura 7.10.3(a), las dos muestras no se solapan en absoluto. Los datos tienen *mínima* compatibilidad con H_0 y muestran la evidencia *más fuerte* a favor de H_A y, por tanto, U_s tiene su valor máximo, $U_s = 20$. De forma similar, $U_s = 20$ en la Figura 7.10.3(b). Por otra parte, la situación *más compatible* con H_0 y que muestra una falta de evidencia a favor de H_A es aquella en la que el solapamiento es máximo, como se muestra en la Figura 7.10.3(c). En esta situación, $K_1 = 10$, $K_2 = 10$, y $U_s = 10$.

Todas las demás posibles disposiciones de los datos estarán en algún lugar entre las tres disposiciones que se muestran en la Figura 7.10.3. Aquellas con mucho solapamiento tendrán valores de U_s cercanos a 10, y aquellas con poco solapamiento tendrán valores de U_s cercanos a 20. Por tanto, valores grandes de U_s indican evidencia a favor de la hipótesis de investigación, H_A o, de forma equivalente, la incompatibilidad de los datos con H_0 .

Consideraremos ahora brevemente la distribución nula de U_s e indicaremos cómo se determinan los valores críticos de la Tabla 6. (Recuérdese de la Sección 7.10 que, en cualquier contraste estadístico, la distribución de referencia de los valores críticos es siempre la distribución nula del estadístico de contraste, es decir, su distribución muestral bajo la condición de que H_0 es cierta). Para determinar la distribución nula de U_s , es necesario calcular las probabilidades asociadas con las diversas disposiciones de los datos, suponiendo que todos los Y se extraen realmente de la misma población*. (El método para calcular las probabilidades se describe brevemente en el Apéndice 7.2).

* Al calcular las probabilidades utilizadas en esta sección, se ha asumido que la probabilidad de que existan observaciones emparejadas es despreciable. Esto será cierto para una variable continua que se mide con una precisión alta. Si el número de emparejamientos es grande, debe realizarse una corrección; véase Noether (1967)⁵⁷.

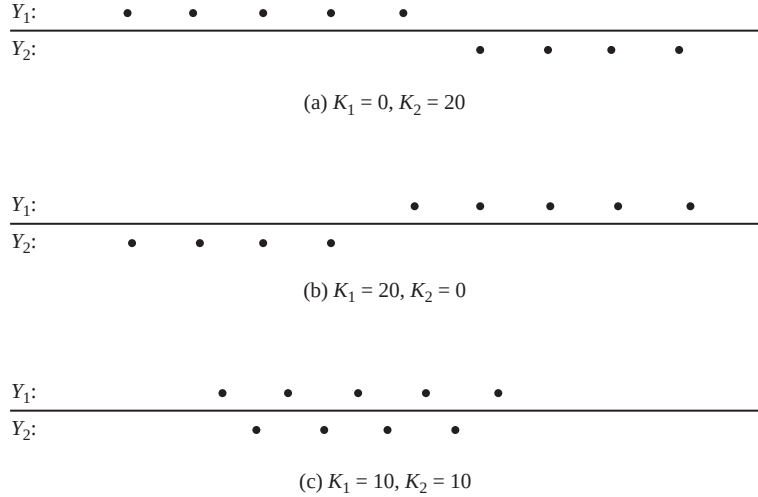


Figura 7.10.3 Tres disposiciones de datos para un test de Wilcoxon-Mann-Whitney

La Figura 7.10.4(a) muestra la distribución nula de K_1 y K_2 para el caso $n = 5$, $n' = 4$. Por ejemplo, se puede demostrar que, si H_0 es cierta, entonces

$$\Pr\{K_1 = 0, K_2 = 20\} = 0,008$$

Esta es la primera probabilidad que se muestra en la Figura 7.10.4(a). Nótese que la Figura 7.10.4(a) es aproximadamente análoga a una distribución t . Valores grandes de K_1 (cola de la derecha) representan evidencia de que los valores de $\bar{Y}_1$ tienden a ser mayores que los valores de $\bar{Y}_2$ y valores grandes de K_2 (cola de la izquierda) representan evidencia de que los valores de $\bar{Y}_2$ tienden a ser mayores que los valores de $\bar{Y}_1$.

La Figura 7.10.4(b) muestra la distribución nula de U_s , que se obtiene directamente de la distribución de la Figura 7.10.4(a). Por ejemplo, si H_0 es cierta, entonces

$$\Pr\{K_1 = 0, K_2 = 0\} = 0,008$$

y

$$\Pr\{K_1 = 20, K_2 = 0\} = 0,008$$

de forma que

$$\Pr\{U_s = 20\} = 0,008 + 0,008 = 0,016$$

que es la probabilidad más a la derecha representada en la Figura 7.10.4(b). Por tanto, ambas colas de la distribución K se han «doblado» sobre la cola superior de la distribución U . Por ejemplo, el área sombreada de una cola en la Figura 7.10.4(b) es igual al área sombreada de las dos colas de la Figura 7.10.4(a).

Los P valores del test de Wilcoxon-Mann-Whitney son áreas de la cola superior de la distribución de U_s . Por ejemplo, se puede demostrar que el área sombreada en azul de la Figura 7.10.4(b) es igual a 0,064. Esto significa que si es cierta, entonces

$$\Pr\{U_s \geq 18\} = 0,064$$

Por tanto, un conjunto de datos para $U_s = 18$ tendría un P valor asociado de 0,064 (suponiendo una H_A no direccional).

Los valores de la Tabla 6 se han determinado a partir de la distribución nula de U_s . Debido a que la distribución de U_s es discreta, solo son posibles unos cuantos P valores para unos tamaños de la muestra dados n_1 y n_2 . La Tabla 6 muestra valores seleccionados de U_s en negrita, con los P valores dados en cursiva. Por ejemplo, si los tamaños muestrales son 5 y 4, entonces un valor de U_s de 17 da un P valor de 0,111, un valor de U_s de 18 da un P valor de 0,064 y un valor de U_s de 19 da un P valor de 0,032. Por tanto, para obtener evidencia significativa con un nivel de significación $\alpha = 0,05$ se requiere

un valor del estadístico de contraste (U_s) de 19. El mínimo P valor posible cuando los tamaños muestrales son de 5 y 4 es 0,016, cuando $U_s = 20$, lo que significa que no se puede obtener evidencia significativa con un nivel de $\alpha = 0,01$ con un test no direccional.

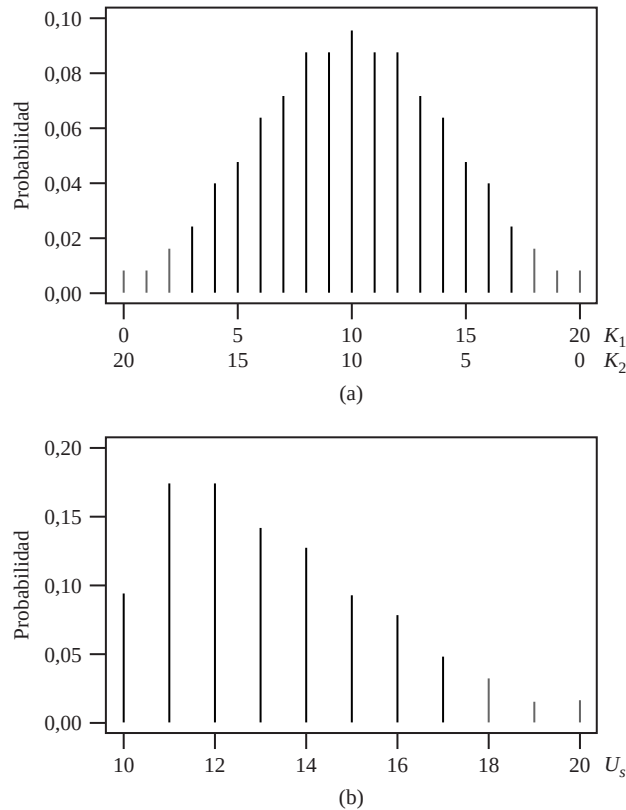


Figura 7.10.4 Distribuciones nulas del test de Wilcoxon-Mann-Whitney cuando $n = 5$, $n' = 4$. (a) Distribución nula de K_1 y K_2 ; (b) distribución nula de U_s . El sombreado corresponde al P valor cuando $U_s = 18$

Condiciones de uso del test de Wilcoxon-Mann-Whitney

Para que el test de Wilcoxon-Mann-Whitney sea aplicable, debe ser razonable considerar los datos como muestras aleatorias de sus respectivas poblaciones, con observaciones independientes dentro de cada muestra y las dos muestras independientes entre sí. Bajo estas condiciones, el test de Wilcoxon-Mann-Whitney es válido independientemente de la forma de las distribuciones poblacionales, suponiendo que la variable Y observada es continua⁵⁸.

Los valores críticos dados en la Tabla 6 se han calculado suponiendo que no hay emparejamientos. Si los datos contienen solo unos pocos emparejamientos, entonces los P valores son aproximadamente correctos*.

El test de Wilcoxon-Mann-Whitney frente al test t y al test de aleatorización

El test de Wilcoxon-Mann-Whitney y el test t intentan responder a la misma pregunta básica: ¿difieren las localizaciones de las dos poblaciones? o ¿tiende una población a tener valores mayores (o menores) que la otra? Pero tratan los datos de forma muy diferentes. A diferencia del test t , el test de Wilcoxon-Mann-Whitney no utiliza los valores reales de

* Realmente, el test de Wilcoxon-Mann-Whitney no está restringido a variables continuas. Se puede aplicar a cualquier variable ordinal. Sin embargo, si Y es discreta o categórica, entonces los datos pueden contener muchos emparejamientos, y el test no deberá usarse sin la modificación apropiada de los valores críticos.

los Y , sino solo sus posiciones relativas en una escala ordenada. Esta es la fuerza y la debilidad del test de Wilcoxon-Mann-Whitney. Por otra parte, el test es independiente de la distribución, porque la distribución nula de U_s solo se relaciona con las diferentes ordenaciones de los Y y, por lo tanto, no depende de la forma de la distribución poblacional. No obstante, el test de Wilcoxon-Mann-Whitney puede ser ineficiente. Puede perder potencia porque no usa toda la información de los datos. Esta ineficiencia es especialmente evidente en el caso de muestras pequeñas.

El test de aleatorización es similar en espíritu al test de Wilcoxon-Mann-Whitney en el sentido de que no depende de la normalidad, aunque la potencia del test de aleatorización es a menudo similar a la del test t . Realizar un test de aleatorización puede ser difícil, y esta es la razón principal por la que los tests de aleatorización no se utilizaron más ampliamente hasta que la potencia computacional llegó a ser algo común.

Ninguno de los competidores (el test de aleatorización, el test t o el test de Wilcoxon-Mann-Whitney) es claramente superior a los otros. Si las distribuciones poblacionales no son aproximadamente normales, el test t podría incluso no ser válido. Además, el test de Wilcoxon-Mann-Whitney puede ser mucho más potente que el test t , especialmente si las distribuciones poblacionales son altamente sesgadas. Si las distribuciones poblacionales son aproximadamente normales con desviaciones típicas iguales, entonces el test t es el mejor, pero sus propiedades son similares a las del test de aleatorización. Para tamaños de muestra moderados, el test de Wilcoxon-Mann-Whitney puede ser casi tan potente como el test t ⁵⁹.

Existe un procedimiento basado en el intervalo de confianza para las medias poblacionales que se asocia con el test de Wilcoxon-Mann-Whitney de la misma forma que el intervalo de confianza para $(\mu_1 - \mu_2)$ se asocia con el test t . El procedimiento está más allá del alcance de este libro.

Ejercicios 7.10.1-7.10.9

7.10.1 Considere dos muestras de tamaños $n_1 = 5$ y $n_2 = 7$. Utilice la Tabla 6 para obtener el P valor, suponiendo que H_A es no direccional y que

- (a) $U_s = 26$
- (b) $U_s = 30$
- (c) $U_s = 35$

7.10.2 Considere dos muestras de tamaños $n_1 = 4$ y $n_2 = 8$. Utilice la Tabla 6 para obtener el P valor, suponiendo que H_A es no direccional y que

- (a) $U_s = 25$
- (b) $U_s = 31$
- (c) $U_s = 32$

7.10.3 En un estudio farmacológico, unos investigadores midieron la concentración química de dopamina en el cerebro en seis ratas expuestas a tolueno y seis ratas de control. (Es el mismo estudio descrito en el Ejemplo 7.2.1). Las concentraciones en la región del striatum del cerebro son las que se muestran en la tabla⁴.

Dopamina (ng/gm)	
Tolueno	Control
3.420	1.820
2.314	1.843
1.911	1.397
2.464	1.803
2.781	2.539
2.803	1.990

- (a) Utilice un test de Wilcoxon-Mann-Whitney para comparar los tratamientos con $\alpha = 0,05$. Utilice una alternativa no direccional.

- (b) Proceda como en el apartado (a), pero con la hipótesis alternativa de que el tolueno aumenta la concentración de dopamina.

7.10.4 En un estudio sobre la hipnosis, se observaron los patrones de respiración de un grupo experimental de sujetos y de un grupo de control. Las medidas de ventilación total (litros de aire por minuto por metro cuadrado de área corporal) se muestran en la tabla que se presenta a continuación⁶⁰ (son los mismos datos que se presentaron en el Ejercicio 7.5.6). Utilice un test de Wilcoxon-Mann-Whitney para comparar los dos grupos con $\alpha = 0,10$. Utilice una alternativa no direccional.

Experimental	Control
5,32	4,50
5,60	4,78
5,74	4,79
6,06	4,86
6,32	5,41
6,34	5,70
6,79	6,08
7,18	6,21

7.10.5 En un experimento para comparar los efectos de dos condiciones de crecimiento diferentes sobre las alturas de crisantemos criados en un invernadero, se encontró que todas las plantas que crecieron bajo la condición 1 eran más altas que cualquiera de las que crecieron bajo la condición 2 (es decir, las dos distribuciones de las alturas no se solapan).

Calcule el valor de U_s y obtenga el P valor si el número de plantas de cada grupo fue

- (a) 3
(b) 4
(c) 5

(Asuma que H_A es no direccional).

7.10.6 En un estudio sobre el comportamiento de autolimpieza de la mosca de la fruta *Drosophila melanogaster*, una única mosca experimental se observó durante tres minutos mientras estaba en una cámara con otras 10 moscas del mismo sexo. El observador apuntó el tiempo de cada episodio («racha») de autolimpieza de la mosca experimental. Este experimento se repitió 15 veces con moscas macho y 15 veces con moscas hembra (con moscas diferentes cada vez). Una cuestión de interés era si había una diferencia con el sexo en el comportamiento de autolimpieza. Los tiempos de autolimpieza observados (tiempo medio por racha, en segundos) fueron los siguientes⁶¹:

Macho: 1,2, 1,2, 1,3, 1,9, 1,9, 2,0, 2,1, 2,2,
2,2, 2,3, 2,3, 2,4, 2,7, 2,9, 3,3

$$\bar{y} = 2,127 \quad s = 0,5936$$

Hembra: 2,0, 2,2, 2,4, 2,4, 2,4, 2,8, 2,8, 2,8,
2,9, 3,2, 3,7, 4,0, 5,4, 10,7, 11,7

$$\bar{y} = 4,093 \quad s = 3,014$$

- (a) Para estos datos, el valor del estadístico de Wilcoxon-Mann-Whitney es $U_s = 189,5$. Utilice un test de Wilcoxon-Mann-Whitney para investigar la diferencia con el sexo en el comportamiento de autolimpieza. Considere la hipótesis H_A no direccional y $\alpha = 0,01$.
- (b) Para estos datos, el error típico de $(\bar{Y}_1 - \bar{Y}_2)$ es $ET \approx 0,7933$ segundos. Utilice un test t para investigar la diferencia con el sexo en el comportamiento de autolimpieza. Considere la hipótesis H_A no direccional y $\alpha = 0,01$.
- (c) ¿Qué condición se requiere para la validez del test t pero no para la validez del test de Wilcoxon-Mann-Whitney? ¿Qué característica o características de los datos sugieren que esta condición puede no cumplirse en este caso?
- (d) Verifique el valor de U_s dado en el apartado (a).

7.10.7 Para probar sustancias potencialmente cancerígenas habitualmente se aplican sobre la piel de un ratón. Surge la cuestión de si un ratón podría adquirir una dosis adicional de la sustancia lamiendo o mordiendo a sus compañeros de jaula. Para responder a esta pregunta, se aplicó el compuesto benzopireno sobre la espalda de 10 ratones. Cinco se enjaularon individualmente y otros cinco se encerraron en una sola jaula. Después de 48 horas, se determinó la concentración del compuesto en el tejido estomacal de cada ratón. Los resultados (nmol/g) fueron los siguientes⁶²:

Enjaulados individualmente	Enjaulados en grupo
3,3	3,9
2,4	4,1
2,5	4,8
3,3	3,9
2,4	3,4

- (a) Utilice un test de Wilcoxon-Mann-Whitney para comparar las dos distribuciones con $\alpha = 0,01$. Utilice como hipótesis alternativa que las concentraciones de benzopireno tienden a ser mayores en los ratones enjaulados en grupo que en los ratones enjaulados individualmente.
- (b) ¿Por qué es válida en este caso una alternativa direccional?

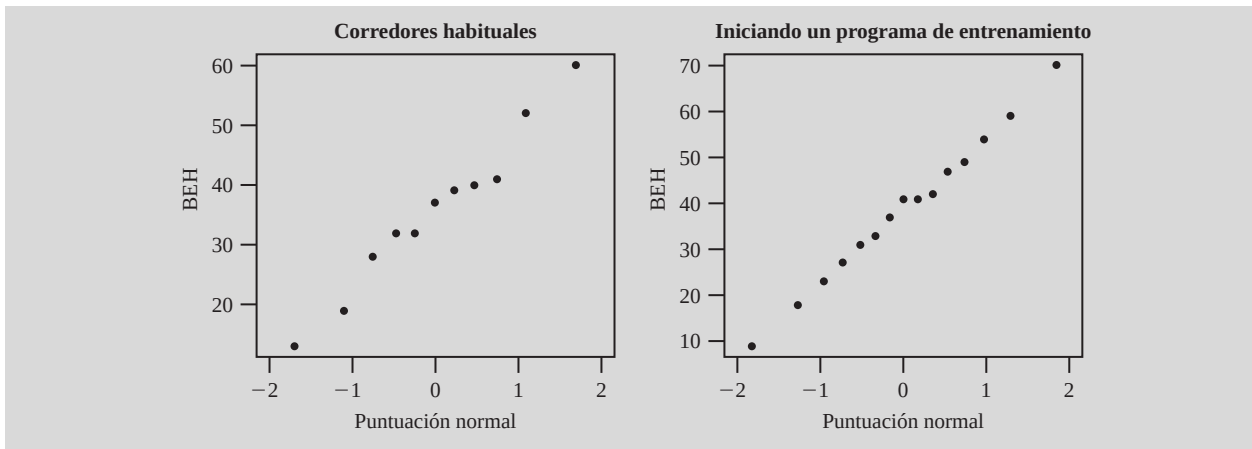
7.10.8 La beta-endorfina humana (BEH) es una hormona segregada por la glándula pituitaria bajo condiciones de estrés. Un ejercicio psicológico midió la concentración de BEH en la sangre en condiciones de reposo (sin estrés) en dos grupos de hombres: el grupo 1 consistía en 11 hombres que corrieron regularmente durante algún tiempo y el grupo 2 consistía en 15 hombres que acababan de entrar en un programa de ejercicio físico. Los resultados se muestran en la siguiente tabla⁶³.

Corredores habituales	Empezando programa de entrenamiento
39 40 32 60	70 47 54 27 31
19 52 41 32	42 37 41 9 18
13 37 28	33 23 49 41 59

Utilice un test de Wilcoxon-Mann-Whitney para comparar las dos distribuciones con $\alpha = 0,10$. Utilice una alternativa no direccional.

Se presentan a continuación gráficas de probabilidad normal de los datos de BEH:

- (a) Utilizando las gráficas para fundamentar su respuesta, ¿hay evidencia de falta de normalidad en alguna de las muestras?
- (b) Considerando su respuesta al apartado (a) y las gráficas anteriores, ¿debería concluirse que los datos tienen en verdad distribución normal? Explique su respuesta.
- (c) Si los datos están en verdad distribuidos normalmente, explique en el contexto de este problema cuál podría ser el inconveniente para utilizar el test de Wilcoxon-Mann-Whitney frente al test t de dos muestras para analizar estos datos.
- (d) Si los datos no están normalmente distribuidos, explique en el contexto de este problema cuál podría ser el inconveniente para utilizar el test t de dos muestras frente al test de Wilcoxon-Mann-Whitney para analizar estos datos.
- (e) Considerando sus respuestas anteriores, argumente qué test se debería utilizar con estos datos. Nótese que hay más de una respuesta correcta.



7.11 Perspectiva

En este capítulo hemos presentado varias técnicas (intervalos de confianza y contrastes de hipótesis) para comparar dos muestras independientes cuando la variable observada cuantitativa. En los capítulos posteriores presentaremos intervalos de confianza y contrastes de hipótesis que son aplicables en varias otras situaciones. Antes de proceder, haremos una pausa para reconsiderar los métodos de este capítulo.

Una suposición implícita

Al presentar los tests de este capítulo (el test t y el test de Wilcoxon-Mann-Whitney) hemos hecho una suposición implícita, que traeremos a la luz a continuación. Al interpretar la comparación de dos distribuciones, hemos supuesto que la relación entre las dos distribuciones es relativamente simple, es decir, que si las distribuciones son diferentes, entonces una de las dos variables tiene una tendencia fuera a ser mayor que la otra. Por ejemplo, supongamos que estamos comparando los efectos de dos dietas en la ganancia de peso de ratones, con

$$\begin{aligned} Y_1 &= \text{Ganancia de peso de ratones con la dieta 1} \\ Y_2 &= \text{Ganancia de peso de ratones con la dieta 2} \end{aligned}$$

Nuestra suposición implícita ha sido que, si las dos dietas son totalmente diferentes, entonces esta diferencia es en la misma dirección para cada ratón individual. Para apreciar el significado de este supuesto, supongamos las distribuciones que se muestran en la Figura 7.11.1. En este caso, incluso aunque la ganancia media de peso es mayor en la dieta 1, sería una sobresimplificación decir que los ratones tienden a ganar más peso con la dieta 1 que con la dieta 2. Aparentemente *algunos* ratones ganan *menos* con la dieta 1. Situaciones paradójicas de este tipo ocurren ocasionalmente, y el análisis simple que se realiza con el test t y el test de Wilcoxon-Mann-Whitney puede ser inadecuado.

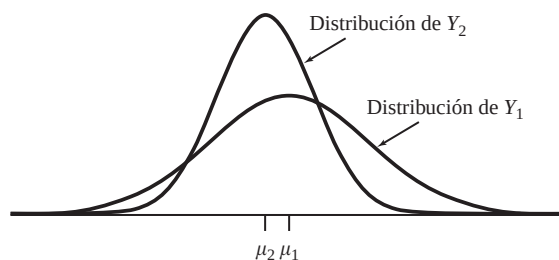


Figura 7.11.1 Distribución de la ganancia de peso con dos dietas

Es relativamente fácil comparar dos distribuciones que tienen la misma forma general y desviaciones típicas similares. Sin embargo, si alguna de las formas o las DT de las dos distribuciones son muy diferentes entre sí, entonces hacer una comparación que tenga sentido de dichas distribuciones es difícil. En particular, una comparación de las dos medias puede no ser apropiada.

Qué método utilizar en cada momento

Si estamos comparando muestras de dos poblaciones distribuidas normalmente, se puede usar un test t para inferir si las medias poblacionales son diferentes y se puede utilizar un intervalo de confianza para estimar en cuánto podrían diferir dichas medias poblacionales, si lo hacen. Un intervalo de confianza proporciona en general más información que un test, ya que el test está restringido a una pregunta puntual («¿Podría la diferencia entre las muestras ser atribuida razonablemente al azar?»), mientras que el intervalo de confianza realiza una pregunta más amplia («¿en cuánto es mayor μ_1 que μ_2 ?»).

Tanto el intervalo de confianza como el test t dependen de la condición de que las poblaciones estén distribuidas normalmente. Si esta condición no se cumple, entonces una transformación podría hacer que las distribuciones fueran aproximadamente normales antes de continuar. Si, a pesar de considerar las transformaciones, la condición de normalidad es cuestionable, entonces se puede usar el test de Wilcoxon-Mann-Whitney. (De hecho, el test de Wilcoxon-Mann-Whitney se puede utilizar si los datos son normales, aunque es menos potente que el test t). Cuando exista duda, una buena forma de proceder es realizar tanto un test t como un test de Wilcoxon-Mann-Whitney. Si los dos tests dan resultados similares y claros (es decir, si los P valores de los tests son similares y ambos son considerablemente mayores que α o ambos son considerablemente menores que α), entonces podemos tener confianza en la conclusión. Sin embargo, si un test t da un P valor algo mayor que α y el otro da un P valor menor que α , entonces deberíamos declarar que los tests no permiten extraer conclusiones.

Algunas veces un *outlier* estará presente en un conjunto de datos, poniendo en cuestión el resultado de un test t . No es legítimo simplemente ignorar el *outlier*. Un procedimiento razonable es llevar a cabo el análisis con el *outlier* incluido y luego eliminar el *outlier* y repetir el análisis. Si la conclusión no cambia cuando el *outlier* está ausente, entonces podemos estar seguros de nosotros mismos que una sola observación no está teniendo excesiva influencia en las conclusiones que extraemos a partir de los datos. Si la conclusión cambia cuando el *outlier* está ausente, entonces no podemos estar seguros de las conclusiones que extraemos. Por ejemplo, si el P valor de un test es pequeño con el *outlier* presente pero grande cuando el *outlier* está ausente, entonces podríamos decir «existe evidencia de que las poblaciones son diferentes entre sí, pero esta evidencia es debida principalmente a una sola observación». Tal afirmación avisa al lector de que no deberían considerarse mucho las diferencias observadas entre las muestras.

Comparación de variabilidad

Sucede algunas veces que es la variabilidad de Y , y no la de su valor medio, la que es de interés principal. Por ejemplo, al comparar dos técnicas diferentes de laboratorio para medir la concentración de la enzima, el investigador podría desear saber principalmente si una de las técnicas es más precisa que la otra, es decir, si su distribución del error de medida tiene una desviación típica menor. Existen técnicas disponibles para contrastar la hipótesis $H_0: \sigma_1 = \sigma_2$ y para usar un intervalo de confianza para comparar σ_1 y σ_2 . La mayor parte de estas técnicas son muy sensibles a la condición de que la distribución subyacente sea normal, lo que limita su uso en la práctica. La implementación de estas técnicas está más allá del alcance de este libro.

Ejercicios suplementarios 7.S.1-7.S.30

(Nota: los ejercicios precedidos por un asterisco se refieren a secciones opcionales). (a)

Las respuestas a las preguntas sobre contrastes de hipótesis deben incluir una afirmación de la conclusión en el contexto del ejercicio. (Véanse los Ejemplos 7.2.4 y 7.2.5).

7.S.1 Para cada una de las siguientes parejas de muestras, calcule el error típico de $(\bar{Y}_1 - \bar{Y}_2)$.

	Muestra 1	Muestra 2
n	12	13
$\bar{y}$	42	47
s	9,6	10,2

(b)

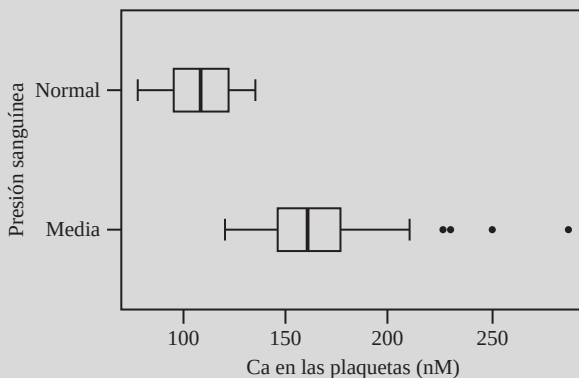
	Muestra 1	Muestra 2
n	22	19
$\bar{y}$	112	126
s	2,7	1,9

(c)

	Muestra 1	Muestra 2
n	5	7
$\bar{y}$	14	16
ET	1,2	1,4

7.5.2 Para investigar la relación entre el calcio intracelular y la presión sanguínea, unos investigadores midieron la concentración de calcio libre en las plaquetas de la sangre de 38 personas con presión sanguínea normal y 45 personas con presión sanguínea alta. Los resultados se muestran en la tabla siguiente y las distribuciones se muestran en los diagramas de caja⁶⁴. Utilice el test t para comparar las medias. Use $\alpha = 0,05$ y considere como H_A la hipótesis no direccional. [Nota: la fórmula (6.7.1) da un valor de 67.5 gl].

	Calcio en las plaquetas (nM)		
Presión sanguínea	n	Media	DT
Normal	38	107,9	16,1
Alta	45	168,2	31,7



7.5.3 En referencia al Ejercicio 7.5.2, construya un intervalo de confianza del 95 % para la diferencia entre las medias poblacionales.

7.5.4 En referencia al Ejercicio 7.5.2, el diagrama de caja para el grupo de presión sanguínea alta está sesgado hacia la derecha e incluye *outliers*. ¿Significa esto que el test t no es válido para estos datos? ¿Por qué o por qué no?

7.5.5 En un estudio de métodos para producir leche de oveja para su uso en la fabricación de queso, se asignaron alea-

toriamente ovejas a un método de ordeño mecánico y a un método manual. El investigador sospechaba que el método mecánico podría irritar las ubres y, por tanto, producir una concentración más alta de células somáticas en la leche. Los datos que se presentan a continuación indican el número medio de células somáticas de cada animal⁶⁵.

	Cuenta somática ($10^{-3} \times$ células/ml)	
	Ordeño mecánico	Ordeño manual
	2.966	186
	269	107
	59	65
	1.887	126
	3.452	123
	189	164
	93	408
	618	324
	130	548
	2.493	139
n	10	10
Media	1.215,6	219,0
DT	1.342,9	156,2

- (a) ¿Dan soporte los datos a la sospecha del investigador? Utilice un test t contra una alternativa direccional con $\alpha = 0,05$. El error típico de $(\bar{Y}_1 - \bar{Y}_2)$ es ET = 427,54 y la fórmula (6.7.1) da como resultado 9,2 gl.
- (b) ¿Dan soporte los datos a las sospechas del investigador? Utilice un test de Wilcoxon-Mann-Whitney contra una alternativa direccional con $\alpha = 0,05$. (El valor del estadístico de Wilcoxon-Mann-Whitney es $U_s = 69$). Compare con el resultado del apartado (a).
- (c) ¿Que condiciones se requieren para la validez del test t pero no para la validez del test de Wilcoxon-Mann-Whitney? ¿Qué característica de los datos arroja dudas sobre esta condición?
- (d) Verifique el valor de U_s dado en el apartado (b).

7.5.6 Un fisiólogo de plantas realizó un experimento para determinar si las tensiones mecánicas podían retrasar el crecimiento de las plantas de soja. Se asignaron aleatoriamente plantas jóvenes a dos grupos con 13 plantas cada uno. Las plantas de un grupo se agitaron mecánicamente durante 20 minutos dos veces al día, mientras que las plantas del otro grupo no se agitaron. Después de 16 días de crecimiento, se midió la longitud total del tallo (cm) de cada planta, y los resultados se muestran en la tabla siguiente⁶⁶.

Utilice un test t para comparar los tratamientos con $\alpha = 0,01$. Tome la hipótesis alternativa que la tensión tiende

a retrasar el crecimiento. [Nota: la fórmula (6.7.1) da un resultado de 23 gl].

	Control	Tensión
n	13	13
$\bar{y}$	30,59	27,78
s	2,13	1,73

7.S.7 En referencia al Ejercicio 7.S.6, construya un intervalo de confianza del 95 % para la reducción de la media poblacional en la longitud del tallo. ¿El intervalo de confianza indica si el efecto de la tensión es «importante para la horticultura», si «importante para la horticultura» se define como una reducción en la media poblacional en la longitud del tallo de al menos

- (a) 1 cm
- (b) 2 cm
- (c) 5 cm

7.S.8 En referencia al Ejercicio 7.S.6, en la tabla siguiente se muestran las observaciones (centímetros) en orden creciente. Compare los tratamientos utilizando un test de Wilcoxon-Mann-Whitney con $\alpha = 0,01$. Considere la hipótesis alternativa como que la tensión tiende a retrasar el crecimiento.

Control	Tensión
25,2	24,7
29,5	25,7
30,1	26,5
30,1	27,0
30,2	27,1
30,2	27,2
30,3	27,3
30,6	27,7
31,1	28,7
31,2	28,9
31,4	29,7
33,5	30,0
34,3	30,6

7.S.9 Una medida del impacto de la polución en un río es la diversidad de especies en su planicie aluvial. En un estudio se compararon dos ríos, el Black River y el Vermilion River. Se muestrearon parcelas de 50 m × 20 m a lo largo de cada río y se contó el número de especies de árboles en cada parcela. La tabla siguiente contiene los datos⁶⁷.

Vermilion River	Black River
9 9 16 13 12	13 10 6 9
13 13 13 8 11	10 7 6 18
9 9 10	6

Se consideraba que el Black River estaba más contaminado que el Vermilion River y, por tanto, se esperaba que hubiera una biodiversidad menor a lo largo del Black River. Realice un test de Wilcoxon-Mann-Whitney con $\alpha = 0,10$, con la hipótesis nula de que las poblaciones de las que se extrajeron las dos muestras tienen la misma biodiversidad (distribución de especies de árboles por parcela) frente a una alternativa direccional apropiada.

7.S.10 Un biólogo de la evolución extrajo los ovocitos (células del óvulo en desarrollo) de los ovarios de 24 ranas (*Xenopus laevis*). Se determinó el pH de cada uno de los ovocitos de cada rana. Además, cada rana fue clasificada de acuerdo a su respuesta a un cierto estímulo con hormona progesterona. Los valores del pH fueron los siguientes⁶⁸:

Respuesta positiva:

7,06, 7,18, 7,30, 7,30, 7,31, 7,32, 7,33, 7,34, 7,36, 7,36, 7,40, 7,41, 7,43, 7,48, 7,49, 7,53, 7,55, 7,57

Sin respuesta:

7,55, 7,70, 7,73, 7,75, 7,75, 7,77

Investigue la relación del pH de los ovocitos con la respuesta a la progesterona utilizando un test de Wilcoxon-Mann-Whitney con $\alpha = 0,05$. Utilice una alternativa no direccional.

7.S.11 En referencia al Ejercicio 7.S.10, la tabla siguiente presenta los estadísticos resumen de las medidas del pH. Investigue la relación del pH de los ovocitos con la respuesta a progesterona utilizando un test t con $\alpha = 0,05$. Utilice una alternativa no direccional. [Nota: la fórmula (6.7.1) da un resultado de 14,1 gl].

	Respuesta positiva	Sin respuesta
n	18	6
$\bar{y}$	7,373	7,708
s	0,129	0,081

7.S.12 Una nueva dieta propuesta para el ganado es menos costosa que la dieta estándar. Los proponentes de la nueva dieta han realizado un estudio comparativo en el que un grupo de ganado fue alimentado con la nueva dieta y otro grupo fue alimentado con la estándar. Encontraron que las ganancias medias de pesos los dos grupos no eran diferentes de forma estadísticamente significativa con nivel de significación del 5 %, y dijeron que este resultado sustentaba la afirmación de que la nueva dieta más barata era tan buena (con respecto a la ganancia de peso) como la dieta estándar. Comente esta afirmación.

***7.S.13** En referencia al Ejercicio 7.S.12, suponga que descubre que el estudio utilizó 25 animales para cada una de las dos dietas, y que el coeficiente de variación de la ganancia de peso bajo las condiciones del estudio era aproximadamente del 20 %. Utilizando esta información adicional, amplíe su

comentario a la afirmación de los proponentes, indicando como sería de probable que un estudio detectara una deficiencia del 10 % en la ganancia de peso de la dieta más barata (utilizando un test de dos colas con nivel de significación del 5 %).

7.S.14 En un estudio sobre la pérdida auditiva, se descubrieron tumores del saco endolinfático (TSEL) en 13 pacientes. Estos 13 pacientes tenían un total de 15 tumores (es decir, la mayoría de los pacientes tenían un único tumor, pero dos de los pacientes tenían dos tumores cada uno). Diez de los tumores se asociaron con la pérdida de la función auditiva en un oído, pero en cinco de los oídos con tumores el paciente no tuvo pérdida auditiva⁶⁹. Una pregunta natural es si la pérdida auditiva es más probable con tumores grandes que con tumores pequeños. Por tanto, se midió el tamaño de los tumores. Suponga que se conocen las medias y las desviaciones típicas muestrales y que se está considerando una comparación del tamaño medio del tumor (pérdida de audición frente a no pérdida de audición).

- (a) Explique por qué no es apropiado aquí un test t para comparar los tamaños medios de los tumores.
 (b) Si se dispone de los datos iniciales, ¿se podría utilizar un test de Wilcoxon-Mann-Whitney?

7.S.15 (Ejercicio para computador) En una investigación sobre la posible influencia del cromo de la dieta sobre los síntomas de la diabetes, se alimentó a 14 ratas con una dieta baja en cromo y a 10 con una dieta normal. Una variable de respuesta era la actividad de la enzima del hígado GITH, que se midió utilizando una molécula marcada radiactivamente. La tabla que se presenta a continuación muestra los resultados, expresados en miles de unidades por minuto por gramo de hígado⁷⁰. Utilice un test t para comparar las dietas con $\alpha = 0,05$. Utilice una alternativa no direccional. [Nota: la fórmula (6.7.1) da un resultado de 29,1 gl].

Dieta baja en cromo		Dieta normal	
42,3	52,8	53,1	53,6
51,5	51,3	50,7	47,8
53,7	58,5	55,8	61,8
48,0	55,4	55,1	52,6
56,0	38,3	47,5	53,7
55,7	54,1		
54,8	52,1		

7.S.16 (Ejercicio para computador) En referencia al Ejercicio 7.S.15, utilice un test de Wilcoxon-Mann-Whitney para comparar las dietas con $\alpha = 0,05$. Utilice una alternativa no direccional.

7.S.17 (Ejercicio para computador) En referencia al Ejercicio 7.S.15:

- (a) Construya un intervalo de confianza del 95 % para la diferencia de las medias poblacionales.

- (b) Suponga que los investigadores creen que el efecto de la dieta baja en cromo es «no importante» si el desplazamiento medio en la actividad de la GIHT es menor del 15 % (es decir, si la diferencia de las medias poblacionales es menor que aproximadamente 8.000 upm/g). De acuerdo con el intervalo de confianza del apartado (a), ¿permiten los datos fundamentar la conclusión de que la diferencia es «no importante»?

- (c) ¿Cómo respondería a la pregunta del apartado (b) si el criterio fue la de 4.000 en vez de 8.000 upm/g?

7.S.18 (Ejercicio para computador) En un estudio sobre el lagarto *Sceloporus occidentalis*, los investigadores examinaron unos lagartos capturados en el campo para ver si estaban infectados por el parásito de la malaria *Plasmodium*. Para ayudar a evaluar el impacto ecológico de la infección por malaria, los investigadores contrastaron la resistencia de 15 animales infectados y 15 animales no infectados, midiendo la distancia que cada animal podía correr en dos minutos. Las distancias (metros) se muestran en la tabla⁷¹.

Animales infectados		Animales no infectados	
16,4	36,7	22,2	18,4
29,4	28,7	34,8	27,5
37,1	30,2	42,1	45,5
23,0	21,8	32,9	34,0
24,1	37,1	26,4	45,5
24,5	20,3	30,6	24,5
16,4	28,3	32,9	28,7
29,1		37,5	

¿Proporcionan los datos evidencia de que la infección está asociada con la disminución de resistencia? Investigue esta cuestión utilizando:

- (a) Un test t .
 (b) Un test de Wilcoxon-Mann-Whitney

Considere H_A direccional y $\alpha = 0,05$.

7.S.19 En un estudio sobre el efecto de la anfetamina en el consumo de agua, un farmacólogo inyectó anfetamina a cuatro ratas y a otras cuatro les inyectó una solución salina para que sirvieran como controles. Midió la cantidad de agua consumida por cada rata en 24 horas. Se presentan a continuación los resultados, expresados en ml de agua por kg de peso corporal⁷².

Anfetamina	Control
118,4	122,9
124,4	162,1
169,4	184,1
105,3	154,9

- (a) Utilice un test t para comparar los tratamientos con $\alpha = 0,10$. Utilice como hipótesis alternativa que la anfetamina tiende a suprimir el consumo de agua.
- (b) Utilice un test de Wilcoxon-Mann-Whitney para comparar los tratamientos con $\alpha = 0,10$, utilizando la alternativa direccional de que la anfetamina tiende a suprimir el consumo de agua.
- (c) ¿Por qué es importante que algunas de las ratas recibieran inyecciones de solución salina como controles? Es decir, ¿por qué los investigadores no compararon simplemente ratas que habían recibido una inyección de anfetamina con ratas que no habían recibido ninguna inyección?

7.5.20 Algunas veces se administra óxido nítrico a recién nacidos que experimentan fallo respiratorio. En un experimento, se administró óxido nítrico a 114 niños. Este grupo se comparó con un grupo de control de 121 niños. Se apuntó el tiempo de hospitalización (en días) de cada uno de los 235 niños. La media muestral del grupo que recibió óxido nítrico fue de $\bar{y}_1 = 36,4$. La media muestral del grupo de control fue de $\bar{y}_2 = 29,5$. Un intervalo de confianza del 95 % para $\mu_1 - \mu_2$ es $(-2,3, 16,1)$, siendo μ_1 la media poblacional del tiempo de hospitalización de los niños que recibieron óxido nítrico y μ_2 la media poblacional de los niños del grupo de control⁷³. Indique si cada una de las siguientes afirmaciones es verdadera o falsa y explique por qué.

- (a) Tenemos un 95 % de confianza en que μ_1 es mayor que μ_2 , ya que la mayor parte del intervalo de confianza es mayor que cero.
- (b) Tenemos un 95 % de confianza en que la diferencia entre μ_1 y μ_2 está entre $-2,3$ días y $16,1$ días.
- (c) Tenemos un 95 % de confianza en que la diferencia entre $\bar{y}_1$ e $\bar{y}_2$ está entre $-2,3$ días y $16,1$ días.
- (d) El 95 % de los niños que recibió óxido nítrico estuvo hospitalizado más tiempo que el promedio de los niños de control.

7.5.21 Considere el intervalo de confianza para $\mu_1 - \mu_2$ en el Ejercicio 7.5.20: $(-2,3, 16,1)$. Indique si es verdadero o falso: si contrastáramos $H_0: \mu_1 = \mu_2$ frente a $H_A: \mu_1 \neq \mu_2$ utilizando $\alpha = 0,05$, rechazaríamos H_0 .

7.5.22 Unos investigadores estudiaron sujetos que tenían neumonía y los clasificaron en dos grupos: aquellos que habían recibido una terapia médica consistente con las directrices de la American Thoracic Society (ATS) y aquellos que habían recibido una terapia médica no consistente con las directrices de la ATS. Los sujetos del grupo «consistente» fueron en general capaces de volver a trabajar antes que los sujetos del grupo «no consistente». Se aplicó un test de Wilcoxon-Mann-Whitney a los datos⁷⁴. El P valor del test fue 0,04. Indique si cada una de las siguientes afirmaciones es verdadera o falsa y explique el motivo.

- (a) Hay una probabilidad del 4 % de que las distribuciones poblacionales «consistente» y «no consistente» serán realmente las mismas.

- (b) Si las distribuciones poblacionales «consistente» y «no consistente» son realmente las mismas, entonces una diferencia entre las dos muestras tan grande como la diferencia observada por los investigadores solo ocurriría el 4 % de las veces.
- (c) Si se realizara un nuevo estudio que compara las poblaciones «consistente» y «no consistente» hay una probabilidad del 4 % de que H_0 fuera rechazada de nuevo.

7.5.23 Un estudiante apuntó de número de calorías de 56 entrantes (28 vegetarianos y 28 no vegetarianos) servidos en un comedor universitario⁷⁵. La tabla que se presenta a continuación contiene un resumen de los datos. Las gráficas de los datos (que no se presentan aquí) muestran que ambas distribuciones son razonablemente simétricas y con forma de campana. Un intervalo de confianza de 95 % para $\mu_1 - \mu_2$ es $(-27, 85)$. Indique si cada una de las siguientes afirmaciones es verdadera o falsa y explique el motivo.

	n	Media	DT
Vegetariana	28	351	119
No vegetariana	28	322	87

- (a) El 95 % de los datos están entre -27 y 85 calorías.
- (b) Tenemos una confianza del 95 % de que $\mu_1 - \mu_2$ esté entre -27 y 85 calorías.
- (c) El 95 % de las veces $\bar{Y}_1 - \bar{Y}_2$ estará entre -27 y 85 calorías.
- (d) El 95 % de los entrantes vegetarianos tendrán entre 27 calorías menos y 85 calorías más que el promedio de las entrantes no vegetarianas.

7.5.24 En referencia al Ejercicio 7.5.23, indique si es verdadero o falso (y explique el motivo): el 95 % de las veces, cuando se realice un estudio de este tipo, la diferencia en las medias muestrales $(\bar{Y}_1 - \bar{Y}_2)$ estará dentro de un intervalo de $(85 - (-27))$ aproximadamente $\frac{85 - (-27)}{2} = 56$ calorías alrededor de

la diferencia de las medias poblacionales $(\mu_1 - \mu_2)$.

7.5.25 (Ejercicio para computador) Las lianas son formaciones leñosas que crecen en los bosques tropicales. Unos investigadores midieron la abundancia de lianas (en tallos/ha) en diversas parcelas de la región central del Amazonas en Brasil. Las parcelas se clasificaron en dos grupos: las que estaban cerca del límite del bosque (a menos de 100 metros del límite) y las que estaban lejos del límite del bosque. Los datos iniciales se muestran y se resumen en la tabla siguiente⁷⁶.

- (a) Realice gráficas de probabilidad normal de los datos para confirmar que las distribuciones son débilmente sesgadas.
- (b) Realice un test t para comparar los dos tipos de parcelas con $\alpha = 0,05$. Utilice una alternativa no direccional.
- (c) Aplique una transformación logarítmica a los datos y repita los apartados (a) y (b).
- (d) Compare los tests t de los apartados (b) y (c). ¿Qué indican estos resultados sobre tests t con sesgo medio cuando los tamaños de las muestras son bastante grandes?

	<i>n</i>	Media	DT
Cerca	34	438	125
Lejos	34	368	114

Cerca			Lejos		
639	601	600	470	339	384
605	581	555	309	395	393
535	531	466	236	252	407
437	423	380	241	215	427
376	362	350	320	228	445
349	346	337	325	267	451
320	317	310	352	294	493
285	271	265	275	356	502
250	450	441	181	418	540
436	432	420	250	425	590
419	407		266	495	
702	676		338	648	

7.S.26 La androstenediona (andro) es un esteroide del que algunos atletas piensan que aumenta la fuerza. Unos científicos investigaron esta afirmación proporcionando andro a un grupo de hombres y un placebo a otro grupo de hombres de control. Una de las variables medidas en el experimento fue el aumento de fuerza de «levantamiento lateral» de cada sujeto después de cuatro semanas. (Un levantamiento lateral es un tipo de ejercicio de levantamiento de peso). Los datos iniciales se presentan a continuación y se resumen en la tabla⁷⁷.

	<i>n</i>	Media	DT
Control	9	14,4	13,3
Andro	10	20,0	12,5

Control				Andro			
30	10	10	30	0	10	0	10
40	20	30	20	10	40	20	10
10	0			30			

- (a) Realice un test *t* para comparar los dos grupos con $\alpha = 0,05$. Utilice una alternativa no direccional. [Nota: la fórmula (6.7.1) da un resultado de 16,5 gl].

- (b) Antes de realizar el estudio se esperaba que la andro aumentaría la fuerza, lo que significa que se podría haber utilizado una alternativa direccional. Rehaga el análisis del apartado (a) utilizando la alternativa direccional apropiada.

7.S.27 A continuación se presenta la salida de computador de un estudio⁷⁸. Describa el problema y la conclusión basándose en la salida del *software* estadístico.

Y = número de bebidas en los siete días anteriores
T de dos muestras para tratamiento vs. control:

	<i>n</i>	Media	DT
Tratamiento	244	13.62	12.39
Control	238	18.86	13.49

95 % IC para $\mu_1 = \mu_2$: (−5.56, −0.92)

Test T $\mu_1 = \mu_2$ (vs <):

T = −2.74 P = .0031 DF = 474.3

7.S.28 En un controvertido estudio para determinar la efectividad del AZT, un grupo de mujeres embarazadas VIH-positivas se asignaron aleatoriamente para tomar AZT o un placebo. Algunos de los bebés nacidos de estas mujeres fueron VIH-positivos, mientras que otros no⁷⁹.

- (a) ¿Cuál es la variable explicativa?
(b) ¿Cuál es la variable de respuesta?
(c) ¿Cuáles son las unidades experimentales?

7.S.29 Unos pacientes que sufrían fallo respiratorio agudo se asignaron aleatoriamente para ser mantenidos en una posición de decúbito prono (boca abajo) o de decúbito supino (boca arriba). En el grupo de decúbito prono, 21 de los 150 pacientes murieron. En el grupo de decúbito supino, 25 de los 152 pacientes murieron⁸⁰.

- (a) ¿Cuál es la variable explicativa?
(b) ¿Cuál es la variable de respuesta?
(c) ¿Cuáles son las unidades experimentales?

7.S.30 Un estudio sobre la terapia de sustitución hormonal (T.S.H.) en mujeres menopáusicas concluyó que tenían una tasa más reducida de ataques al corazón, y que tenían incluso una reducción mayor en muertes por homicidios y accidentes (dos causas de muerte que no se podrán relacionar con la T.S.H.). Parece que las mujeres con T.S.H. difieren de las otras en muchos otros aspectos de sus vidas (por ejemplo, practican más ejercicio, tienden a ser más saludables y mejor educadas)⁸¹. Utilice el lenguaje de la estadística para comentar lo que dirían estos datos sobre las relaciones entre T.S.H., riesgo de ataque al corazón y variables como ejercicio, salud y educación. Utilice un diagrama esquemático similar al de la Figura 7.4.1 o al de la Figura 7.4.2 para fundamentar su explicación.

Notas

1. Kotler, D. (2000). A comparison of aerobics and modern dance training on health-related fitness in college women. Manuscrito no publicado, Oberlin College.
2. Pappas, T. y Mitchell, C. A. (1985). Effects of seismic stress on the vegetative growth of *Glycine max* (L.) Merr. cv. Wells II. *Plant, Cell and Environment* **8**, 143-148. Reimpreso con permiso de Blackwell Scientific Publications Limited. Datos iniciales cortesía de los autores. El experimento original incluía muchos tratamientos y más de nueve observaciones por grupo. Por simplicidad, aquí solo se presenta un subconjunto de los datos.
3. Datos no publicados cortesía de J. A. Henricksy V. J. K. Liu.
4. Rea, T. M., Nash, J. F., Zabik, J. E., Born, G. S. y Kessler, W. V. (1984). Effects of toluene inhalation on brain biogenic amines in the rat. *Toxicology* **31**, 143-150. Datos iniciales cortesía de J. F. Nashy y J. E. Zabik.
5. Sagan, C. (1977). *The Dragons of Eden*. New York: Ballantine, p. 7.
6. Lemenager, R. P., Nelson, L. A. y Hendrix, K. S. (1980). Influence of cow size and breed type on energy requirements. *Journal of Animal Science* **51**, 566-576. Algunos animales perdieron peso durante los 78 días, de forma que las ganancias medias de peso se basan en valores positivos y negativos.
7. Adaptado de Miyada, V. S. (1978). Uso da levedura seca de destilarias de álcool de cana de açúcar na alimentação de suínos em crescimento e acabamento. Master's thesis, University of Sao Paulo, Brazil.
8. Kalsner, S. y Richards, R. (1984). Coronary arteries of cardiac patients are hyperreactive and contain stores of amines: A mechanism for coronary spasm. *Science* **223**, 1435-1437. Copyright 1984 de la American Association for the Advancement of Science (AAAS).
9. Adaptado de Dybas, H. S. y Lloyd, M. (1962). Isolation by habitat in two synchronized species of periodical cicadas (Homoptera, Cicadidae, *Magicicada*). *Ecology* **43**, 444-459.
10. Namdar, M., Koepfli, P., Grathwohl, R., Siegrist, P. T., Klainguti, M., Schepis, T., Deleloye, R., Wyss, C. A., Gaemperli, O. y Kaufmann, P. A. (2006). Caffeine decreases exercise-induced myocardial flow reserve. *Journal of the American College of Cardiology* **47**, 405-410. Datos iniciales obtenidos de la Figura 1.
11. Bockman, D. E. y Kirby, M. L. (1984). Dependence of thymus development on derivatives of the neural crest. *Science* **223**, 498-500. Copyright 1984 de la AAAS.
12. Tripepi, R. R. y Mitchell, C. A. (1984). Metabolic response of river birch and European birch roots to hypoxia. *Plant Physiology* **76**, 31-35. Datos iniciales cortesía de los autores.
13. Lamke, L. O. y Liljedahl, S. O. (1976). Plasma volume changes after infusion of various plasma expanders. *Resuscitation* **5**, 93-102.
14. Anderson, J. W., Story, L., Sieling, B., Chen, W. J. L., Petro, M. S. y Story, J. (1984). Hypocholesterolemic effects of oat-bran or bean intake for hypercholesterolemic men. *The American Journal of Clinical Nutrition* **40**, 1146-1155.
15. Ahne, A. y Myers, S. (1999). The effect of Miracle Grow on radish growth. Manuscrito no publicado, Oberlin College. Los datos presentados aquí son un subconjunto del conjunto completo. (Las medias y desviaciones típicas del conjunto completo son similares a las del subconjunto presentado aquí. En particular, la media muestral del grupo de control es mayor que la del grupo del fertilizante).
16. Borg, E. (2008). A comparison of *Orconectes rusticus* and *O. sanbornii* weight. Manuscrito no publicado, Oberlin College.
17. Heald, F. (1974). Hematocrit values of youths 12-17 years, United States. *U.S. National Center for Health Statistics, Vital and Health Statistics*, Serie 11, n.º 146. Washington, D.C.: U.S. Department of Health, Education and Welfare. Realmente, los datos se obtuvieron mediante un esquema de muestreo más complicado que el muestreo aleatorio simple.
18. Long, T. F., Murdock, L. L. (1983). Stimulation of blowfly feeding behavior by octopaminergic drugs. *Proceedings of the National Academy of Sciences* **80**, 4159-4163. Datos iniciales cortesía de los autores y de L. C. Sudlow.
19. Yerushalmy, J. (1971). The relationship of parents' cigarette smoking to outcome of pregnancy-implications as to the problem of inferring causation from observed associations. *American Journal of Epidemiology* **93**, 443-456.
20. Gould, S. J. (1981). *The Mismeasure of Man*. New York: Norton, pp. 50ff. Las DT fueron estimadas a partir de los intervalos indicados por Gould.
21. Yerushalmy, J. (1972). Infants with low birth weight born before their mothers started to smoke cigarettes. *American Journal of Obstetrics and Gynecology* **112**, 277-284.
22. Anderson, G. D., Blidner, I. N., McClemon, S. y Sinclair, J. C. (1984). Determinants of size at birth in a Canadian population. *American Journal of Obstetrics and Gynecology* **150**, 236-244.
23. Mochizuki, M., Marno, T., Masuko, K. y Ohtsu, T. (1984). Effects of smoking on fetoplacental-maternal system during pregnancy. *American Journal of Obstetrics and Gynecology* **149**, 413-420.

24. Wainright, R. L. (1983). Change in observed birth weight associated with a change in maternal cigarette smoking. *American Journal of Epidemiology* **117**, 668-675.
25. Moore, R. M., Diamond, E. L. y Cavalieri, R. L. (1988). The relationship of birth weight and intrauterine diagnostic ultrasound exposure. *Obstetrics and Gynecology* **71**, 513-517.
26. Waldenstrom, U., Nilsson, S., Fall, O., Axelsson, O., Eklund, G., Lindeberg, S. y Sjodin, Y. (1988). Effects of routine one-stage ultrasound screening in pregnancy: A randomized clinical trial. *Lancet* (10 sept.), 585-588.
27. National Center for Health Statistics. Los datos se han tomado de la Tabla LCWK9 (www.cdc.gov/nchs/datawh/statab/unpubd/mortabs/lcwk9_10.htm).
28. Cook, L. S., Daling, J. R., Voigt, L. F., de Hart, M. P., Malone, K. E., Stanford, J. L., Weiss, N. S., Brinton, L. A., Gammon, M. D. y Brogan, D. (1997). Characteristics of women with and without breast augmentation. *Journal of the American Medical Association* **277**, 1612-1617.
29. LaCroix, A. Z., Mead, L. A., Liang, K., Thomas, C. B., y Pearson, T. A. (1986). Coffee consumption and the incidence of coronary heart disease. *New England Journal of Medicine* **315**, 977-982.
30. Yerushalmy, J. y Hilleboe, H. E. (1957). Fat in the diet and mortality from heart disease. *New York State Journal of Medicine* **57**, 2343-2354. Reimpreso con permiso. Copyright de la Medical Society of the State of New York.
31. *Cleveland Plain Dealer*, 10 de febrero de 1999, p. 17-A.
32. David, R. J. y Collins, J. W. (1997). Differing birth weight among infants of U.S.-born blacks, African-born blacks, and U.S.-born whites. *New England Journal of Medicine* **337**, 1209-1214. Bajo peso al nacer significa un peso menor que 1.500 g, que los autores denominan en el artículo «muy bajo peso al nacer».
33. Gwilyn, S., Howard, D. P. J., Davies, N. y Willett, K. (2005). Harry Potter casts a spell on accident prone children. *British Medical Journal* **331**, 1505-1506. Los autores advirtieron que los fines de semana «de Harry Potter» tuvieron buen tiempo y que el tiempo no fue muy diferente del de los otros fines de semana.
34. Adaptado de datos proporcionados por cortesía de D. R. Shields y D. K. Colby. Véase Shields, D. R. (1981). The influence of niacin supplementation on growing ruminants and *in vivo* and *in vitro* rumen parameters. Ph. D. thesis, Purdue University.
35. Schall, J. J., Bennett, A. F. y Putnam, R. W. (1982). Lizards infected with malaria: Physiological and behavioral consequences. *Science* **217**, 1057-1059. Copyright 1982 de la AAAS.
36. Agosti, E. y Camerota, G. (1965). Some effects of hypnotic suggestion on respiratory function. *International Journal of Clinical and Experimental Hypnosis* **13**, 149-156.
37. Adaptado de Knight, S. L. y Mitchell, C. A. (1983). Enhancement of lettuce yield by manipulation of light and nitrogen nutrition. *Journal of the American Society for Horticultural Science* **108**, 750-754.
38. Rickard, I. J. (2008). Offspring are lighter at birth and smaller in adulthood when born after a brother versus a sister in humans. *Evolution and Human Behavior* **29**, 196-200.
39. Datos no publicados cortesía de J. L. Wolfson.
40. Datos ficticios pero realistas.
41. Massey, R. L. (2010). A randomized trial of rocking-chair motion on the effect of postoperative ileus duration in patients with cancer recovering from abdominal surgery. *Applied Nursing Research* **23**, 59-64.
42. Shima, J. S. (2001). Recruitment of a coral reef fish: Roles of settlement, habitat, and postsettlement losses. *Ecology* **82**, 2190-2199. Datos iniciales cortesía del autor.
43. Adaptado de Williams, G. Z., Widdowson, G. M. y Penton, J. (1978). Individual character of variation in time-series studies of healthy people. II. Differences in values for clinical chemical analytes in serum among demographic groups, by age and sex. *Clinical Chemistry* **24**, 313-320.
44. Datos ficticios pero realistas. Véase Abraham, S., Johnson, C. L. y Najjar, M. F. (1979). Weight and height of adults 18-74 years of age, United States 1971-74. U.S. National Center for Health Statistics, *Vital and Health Statistics*, Series 11, n.º 211. Washington, D.C.: U.S. Department of Health, Education and Welfare.
45. Ejemplo comunicado por D. A. Holt.
46. Petrie, B. y Segalowitz, S. J. (1980). Use of fetal heart rate, other perinatal and maternal factors as predictors of sex. *Perceptual and Motor Skills* **50**, 871-874. Copyright 1980 de Ammons Scientific, Ltd. Reproducido con permiso de Ammons Scientific, Ltd. via Copyright Clearance Center.
47. Hagerman, A. E. y Nicholson, R. L. (1982). High-performance liquid chromatographic determination of hydroxycinnamic acids in the maize mesocotyl. *Journal of Agricultural and Food Chemistry* **30**, 1098-1102. Copyright 1982 American Chemical Society. Reimpreso con permiso.
48. Ressler, S. (1977) AnthroKids-Anthropometric data of children. Datos tomados del archivo individuals.csv en ovrt.nist.gov/projects/anthrokids/
49. Adaptado de Williams, G. Z., Widdowson, G. M. y Penton, J. (1978). Individual character of variation in time-series studies of healthy people. II. Difference in values for clinical chemical analytes in serum among demographic groups, by age and sex. *Clinical Chemistry* **24**, 313-320. Reimpreso con permiso.

50. Hamill, P. V. V., Johnston, F. E., y Lemeshow, S. (1973). Height and weight of youths 12-17 years, United States. *U.S. National Center for Health Statistics, Vital and Health Statistics, Series 11*, n.º 124. Washington, D.C.: U.S. Department of Health, Education and Welfare.
51. Phelan, S. y Schaffner, A. (2009). NIH grant proposal: Prevention of postpartum weight retention in low-income WIC women.
52. Roberts, J. (1975). Blood pressure of persons 18-74 years, United States, 1971-72. *U.S. National Center for Health Statistics, Vital and Health Statistics, Series 11*, n.º 150. Washington, D.C.: U.S. Department of Health, Education and Welfare. Sin embargo, la distribución de la presión sanguínea sistólica está más sesgada (véase el Ejercicio 5.2.18).
53. Pearson, E. S. y Please, N. W. (1975). Relation between the shape of population distribution and the robustness of four simple tests tatistics. *Biometrika* **62**, 223-241.
54. Mena, E. A., Kossovsky, N., Chu, C. y Hu, C. (1995). Inflammatory intermediates produced by tissues encasing silicone breast implants. *Journal of Investigative Surgery* **8**, 31-42. [Nota: Había dos grupos de control en el estudio. El grupo de control incluido en el análisis es «pacientes que han tenido mamoplastia inversa de aumento» (el grupo «scar» presentado en el artículo). Además, los autores evitaron transformar los datos antes de realizar el test *t*. Por tanto, obtuvieron un *P* valor grande, aunque advirtieron que los dos grupos parecían bastante diferentes].
55. Datos ficticios pero realistas. Basado en datos no publicados proporcionados por Bill Plummer.
56. Fierer, N. (1994). Statistical analysis of soil respiration rates in a light gap and surrounding old-growth forest. Manuscrito no publicado, Oberlin College.
57. Noether, G. E. (1967). *Elements of Nonparametric Statistics*. New York: Wiley.
58. Algunas veces se dice que la validez del test de Mann-Whitney requiere que las dos distribuciones poblacionales tengan la misma forma y se diferencien solo por un desplazamiento. Esto no es correcto. Los cálculos de la Tabla 6 requieren solo que la distribución de la población común (bajo la hipótesis nula) sea continua. Una propiedad posterior, denominada técnicamente *consistencia* del test, requiere que las dos distribuciones del test estén *estocásticamente ordenadas*, que es la forma técnica de decir que una de las variables tiene una tendencia consistente a ser mayor que la otra. De hecho, el título original del artículo de Mann y Whitney es «Sobre un test de si una dedos variables aleatorias es estocásticamente mayor que la otra» (*Annals of Mathematical Statistics* **18**, 1947). En la Sección 7.12 explicamos el requisito de ordenación estocástica, denominándolo «supuesto implícito». (El intervalo de confianza mencionado al final de la Sección 7.10 requiere el supuesto más fuerte de que las distribuciones tengan la misma forma).
59. Zimmerman, D. W. y Zumbo, B. D. (1993). The relative power of parametric and nonparametric statistical methods in G. Keren and C. Lewis (Eds.), *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, pp. 481-517. Hillsdale, N.J.: Lawrence Erlbaum Associates. Los autores usan simulaciones para mostrar que el test de Wilcoxon-Mann-Whitney es más potente que el test *t* en presencia de *outliers*, pero que en ausencia de *outliers*, el test *t* es ligeramente preferible para una variedad de distribuciones poblacionales.
60. Agosti, E. y Camerota, G. (1965). Some effects of hypnotic suggestion on respiratory function. *International Journal of Clinical and Experimental Hypnosis* **13**, 149-156.
61. Connolly, K. (1968). The social facilitation of preening behaviour in *Drosophila melanogaster*. *Animal Behaviour* **16**, 385-391.
62. Datos no publicados cortesía de G. P. Carlson y M. A. Morse.
63. Lobstein, D. D. (1983). A multivariate study of exercise training effects on beta-endorphin and emotionality in psychologically normal, medically healthy men. Ph.D. thesis, Purdue University. Datos iniciales cortesía del autor.
64. Erne, P., Bolli, P., Buerigisser, E. y Buehler, F. R. (1984). Correlation of platelet calcium with blood pressure. *New England Journal of Medicine* **310**, 1084-1088. Reimpreso con permiso del *New England Journal of Medicine*. Estadísticos resumen calculados de datos iniciales proporcionados por cortesía de F. R. Buehler.
65. Adaptado de datos no publicados proporcionados por F. Delgado. El valor extremadamente alto de células somáticas representa probablemente casos de mastitis.
66. Pappas, T. y Mitchell, C. A. (1985). Effects of seismic stress on the vegetative growth of *Glycine max* (L.) Merr. cv. Wells II. *Plant, Cell and Environment* **8**, 143-148. Reimpreso con permiso de Blackwell Scientific Publications Limited. Datos iniciales cortesía de los autores. El experimento original incluía más de dos grupos de tratamiento.
67. Wee, K. (1995). Species diversity in floodplain forests. Manuscrito no publicado, Oberlin College.
68. Cicirelli, M. F., Robinson, K. R. y Smith, L. D. (1983). Internal pH of *Xenopus* oocytes: A study of the mechanism and role of pH changes during meiotic maturation. *Developmental Biology* **100**, 133-146. Datos iniciales cortesía de M. F. Cicirelli.
69. Manski, T. J., et al. (1997). Endolymphatic sac tumors: A source of morbid hearing loss in von Hippel-Lindau disease. *Journal of the American Medical Association* **277**, 1461-1466.
70. Datos no publicados cortesía de J. A. Henricks y V. J. K. Liu.

71. Schall, J. J., Bennett, A. F. y Putnam, R. W. (1982). Lizards infected with malaria: Physiological and behavioral consequences. *Science* **217**, 1057-1059. Copyright 1982 by the AAAS. Datos iniciales cortesía de J. J. Schall.
72. Datos no publicados cortesía de M. B. Nichols y R. P. Maickel.
73. The neonatal inhaled nitric oxide study group (1997). Inhaled nitric oxide in full-term and nearly full-term infants with hypoxic respiratory failure. *New England Journal of Medicine* **336**, 597-604.
74. Gleason, P. P., et al. (1997). Medical outcomes and antimicrobial costs with the use of American Thoracic Society guidelines for outpatients with community-acquired pneumonia. *Journal of the American Medical Association* **278**, 32-39.
75. Hodapp, M. (1998). A Study of CDS Nutrition. Manuscrito no publicado, Oberlin College.
76. Laurance, W. F., Perez-Salicrup, D., Delamonica, P., Fearside, P. M., D'Angelo, S., Jerozolinski, A., Pohl, L. y Lovejoy, T. E. (2001). Rain forest fragmentation and the structure of Amazonian liana communities. *Ecology* **82**, 105-116. Los datos presentados fueron tomados por J. Witmer de la Figura 2 del artículo y pueden no ser completamente exactos.
77. King, D. S., Sharp, R. L., Vukovich, M. D., Brown, G. A., Reifenrath, T. A., Uhl, N. L. y Parsons, K. A. Effect of oral androstenedione on serum testosterone and adaptations to resistance training in young men. *Journal of the American Medical Association* **281**, 2020-2028. Datos iniciales cortesía de los autores. La variable de respuesta que se muestra aquí es el cambio en la «fuerza muscular máxima», que es el máximo peso que el sujeto puede levantar. En el experimento se tomaron otras varias medidas; generalmente mostraron los mismos resultados vistos en los datos de tirón lateral. Un objetivo importante del experimento fue estudiar el efecto del andro en el nivel de testosterona. Los investigadores descubrieron que el andro no tenía efecto en el nivel de testosterona en suero.
78. Fleming, M. F., Barry, K. L., Manwell, L. B., Johnson, K. y London, R. (1997). Brief physician advice for problem alcohol drinkers. *Journal of the American Medical Association* **277**, 1039-1045.
79. Conner, E. M., Sperling, R. S., Gerber, R., Kisalev, P., Scott, G., O'Sullivan, M. J., Van Dyke, R., Bey, M., Shearer, W., Jacobsen, R. L., Jimenez, E., O'Neill, E., Bazin, B., Delfraissy, J.-F., Culname, M., Coombs, R., Elkins, M., More, J., Stratton, P. y Balsley, J. (1994). Reduction of maternal-infant transmission of Human Immunodeficiency Virus Type I with zidovudine treatment. *New England Journal of Medicine* **331**, 1173-1180. Algunas personas piensan que este estudio no se debería haber realizado como un experimento aleatorizado, ya que había razones para pensar que la AZT sería útil para evitar la transferencia del VIH a los bebés, y ya que el VIH causa una dolencia tan seria.
80. Gattinoni, L., Tognoni, G., Pesenti, A., Taccone, P., Mascheroni, D., Labarta, V., Malacrida, R., DiGiulio, P., Fumagalli, R., Pelosi, P., Brazzi, L. y Latini, R. (2001). Effect of prone positioning on the survival of patients with acute respiratory failure. *New England Journal of Medicine* **345**, 568-573.
81. Petitti, D. B., Perlman, J. A. y Sidney, S. (1987). Noncontraceptive estrogens and mortality: Long-term follow-up of women in the Walnut Creek Study. *Obstetrics & Gynecology* **70**, 289-293.

COMPARACIÓN DE DOS MUESTRAS PAREADAS

8

Objetivos

En este capítulo estudiaremos las comparaciones de muestras pareadas. Concretamente:

- demostraremos cómo realizar un test t de muestras pareadas;
- demostraremos cómo construir e interpretar un intervalo de confianza para la media de una diferencia de muestras pareadas;
- presentaremos formas en las que surgen los datos pareados y cómo el emparejamiento puede ser ventajoso;
- consideraremos las condiciones bajo las que un test t de muestras pareadas es válido;
- mostraremos cómo se pueden analizar datos emparejados utilizando el test de signos y el test de rango de signos de Wilcoxon.

8.1 Introducción

En el Capítulo 7 consideramos la comparación de dos muestras independientes cuando la variable de respuesta Y es una variable cuantitativa. En el presente capítulo consideraremos la comparación de dos muestras que no son independientes, sino que están emparejadas. En un **diseño emparejado**, las observaciones (Y_1, Y_2) ocurren en parejas. Las unidades observacionales en una pareja están relacionadas de alguna forma, de manera que tienen más en común entre sí que con los miembros de otra pareja. Presentamos a continuación un ejemplo de diseño pareado.

Ejemplo 8.1.1 Flujo sanguíneo

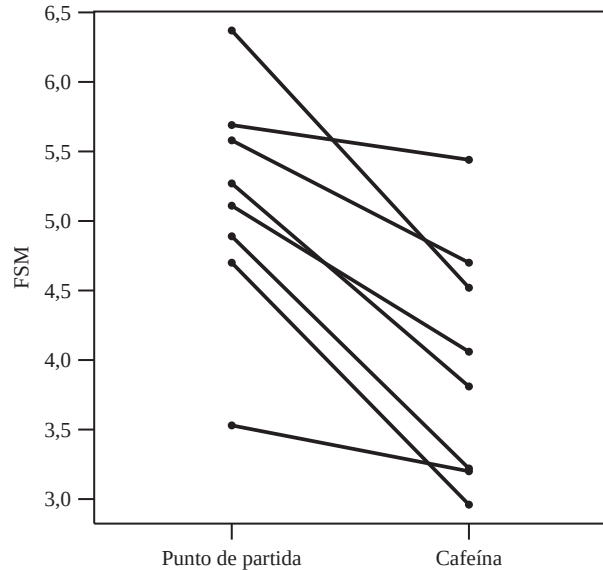
¿Afecta al flujo sanguíneo el beber café, particularmente durante el ejercicio? Unos doctores que estaban estudiando sujetos sanos midieron el flujo de sangre en el miocardio (FSM)* durante el ejercicio en bicicleta antes y después de dar a los sujetos una dosis de cafeína equivalente a beber dos tazas de café. La Tabla 8.1.1 muestra los niveles de FSM antes (punto de partida) y después (cafeína) de que los sujetos tomaran una tableta que contenía 200 mg de cafeína¹. La Figura 8.1.1 muestra diagramas de puntos en paralelo de estos datos, con líneas que conectan las lecturas del punto de partida y de la cafeína de cada sujeto, de forma que es evidente el cambio del «antes» al «después». ■

En el Ejemplo 8.1.1 los datos surgen en parejas. Los datos en la pareja están relacionados en virtud de que son medidas de la misma persona. Un análisis adecuado de los datos debería aprovechar esta emparejamiento. Es decir, podríamos imaginar un experimento en el que algunos sujetos son estudiados después de recibir la cafeína y otros son estudiados

* El FSM se midió mediante imágenes de tomografía por emisión de positrones (TEP) tras administrar a los pacientes agua marcada con oxígeno-15.

Tabla 8.1.1 Flujo de sangre en el miocardio (ml/min/g) de ocho sujetos

Sujeto	FSM	
	Punto de partida y_1	Cafeína y_2
1	6,37	4,52
2	5,69	5,44
3	5,58	4,70
4	5,27	3,81
5	5,11	4,06
6	4,89	3,22
7	4,70	2,96
8	3,53	3,20
Media	5,14	3,99
DT	0,83	0,86

**Figura 8.1.1** Diagramas de puntos de lecturas de FSM antes y después del consumo de cafeína, con rectas conectando las lecturas de cada sujeto

sin haber recibido nunca cafeína. Un experimento como ese proporcionaría dos muestras de datos independientes y podría ser analizado utilizando los métodos del Capítulo 7. Pero el experimento actual utilizó un diseño emparejado. Un flujo de sangre en el miocardio varía de persona a persona, y algunos sujetos tienen niveles altos de FSM tanto antes como después de consumir cafeína, mientras que otros tienen niveles de FSM bajos. Conocer el nivel de FSM en el punto de partida nos dice algo sobre cómo actuará la cafeína sobre el sujeto, y viceversa. Nuestro objetivo es utilizar esta información cuando analizamos los datos.

En la Sección 8.2 mostraremos cómo analizar datos pareados utilizando métodos basados en la distribución t de Student. En las Secciones 8.4 y 8.5 describiremos los tests no paramétricos para datos empareados. Las Secciones 8.3, 8.6 y 8.7 contienen más ejemplos y comentarios del diseño pareado.

8.2 El intervalo de confianza y el test t para muestras pareadas

En esta sección presentaremos el uso de la distribución t de Student para obtener tests e intervalos de confianza para datos emparejados.

Análisis de diferencias

En el Capítulo 7 consideramos cómo analizar datos de dos muestras independientes. Cuando tenemos datos emparejados, simplemente cambiamos nuestro punto de vista: en vez de considerar Y_1 e Y_2 de forma separada, consideraremos la *diferencia* D , definida como

$$D = Y_1 - Y_2$$

Nótese que frecuentemente es natural considerar una diferencia como la variable de respuesta de interés en un estudio. Por ejemplo, si estuviéramos estudiando las velocidades de crecimiento de plantas, podríamos hacer crecer plantas bajo condiciones de control durante un tiempo al comienzo de un estudio y después aplicar un tratamiento durante una semana. Podríamos medir el crecimiento que tiene lugar durante la semana después del tratamiento como $D = Y_1 - Y_2$, siendo Y_1 = altura una semana después de aplicar el tratamiento e Y_2 = altura antes de aplicar el tratamiento*. Algunas veces los

* Los Ejercicios 7.2.11 y 7.2.12 consideran datos «después versus antes».

datos están emparejados de forma que es menos obvia, pero siempre que tengamos datos emparejados, son las diferencias observadas lo que desearemos analizar.

Denominaremos $\bar{D}$ la media de las diferencias muestrales de D . La cantidad $\bar{D}$ está relacionada con las medias muestrales individuales como sigue:

$$\bar{D} = (\bar{Y}_1 - \bar{Y}_2)$$

La relación entre las medias poblacionales es análoga:

$$\mu_D = \mu_1 - \mu_2$$

Por tanto, podemos decir que *la media de la diferencia es igual a la diferencia de las medias*. Debido a esta relación simple, la comparación de dos medias emparejadas se puede realizar concentrándose completamente en los valores de D .

El error típico de $\bar{D}$ es fácil de calcular. Como $\bar{D}$ es simplemente la media de una única muestra, podemos aplicar la fórmula del ET del Capítulo 6 para obtener la siguiente fórmula:

$$ET_{\bar{D}} = \frac{s_D}{\sqrt{n_D}}$$

Siendo s_D la desviación típica de los valores de D y n_D el número de valores de D . El siguiente ejemplo ilustra el cálculo.

Ejemplo 8.2.1 Flujo sanguíneo

La Tabla 8.2.1 muestra los datos de flujo sanguíneo del Ejemplo 8.1.1 y las diferencias d . Nótese que la media de la diferencia es igual a la diferencia de las medias:

$$\bar{d} = 1,15 = 5,14 - 3,99$$

La Figura 8.2.1 muestra la distribución de las ocho diferencias muestrales.

Tabla 8.2.1 Flujo de sangre en el miocardio (ml/min/g) de ocho sujetos			
Sujeto	FSM		
	Punto de partida y_1	Cafeína y_2	Diferencia $d = y_1 - y_2$
1	6,37	4,52	1,85
2	5,69	5,44	0,25
3	5,58	4,70	0,88
4	5,27	3,81	1,46
5	5,11	4,06	1,05
6	4,89	3,22	1,67
7	4,70	2,96	1,74
8	3,53	3,20	0,33
Media	5,14	3,99	1,15
DT	0,83	0,86	0,63

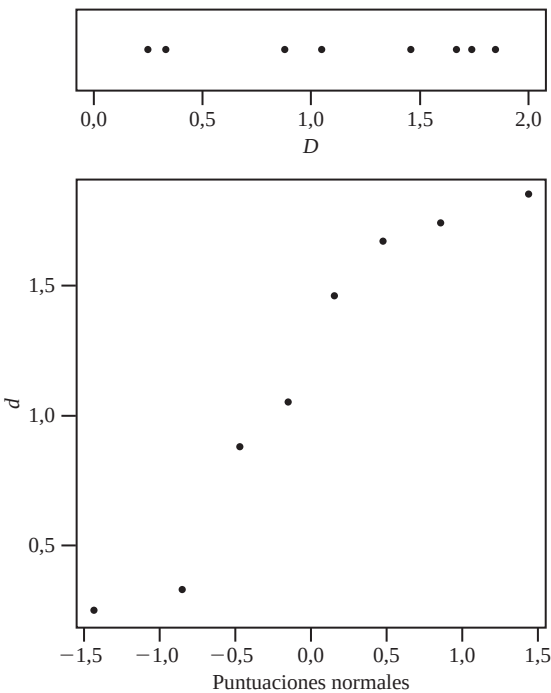


Figura 8.2.1 Diagrama de puntos de lecturas de FSM en el punto de partida y después del consumo de cafeína, junto con una gráfica de probabilidad normal de los datos

Calculamos el error típico de la diferencia media como sigue:

$$\begin{aligned}s_D &= 0,63 \\ n_D &= 8 \\ ET_{\bar{D}} &= \frac{0,63}{\sqrt{8}} = 0,22\end{aligned}$$

Aunque la media de la diferencia es lo mismo que la diferencia de las medias, nótese que el error típico de la diferencia media *no* es la diferencia de los errores típicos de las medias. ■

Intervalo de confianza y contraste de hipótesis

El error típico descrito previamente es la base del **método de análisis t para muestras pareadas**, que puede tomar la forma de un intervalo de confianza o de un contraste de hipótesis.

Un intervalo de confianza del 95 % para μ_D se construye como

$$\bar{d} \pm t_{n_D-1, 0,025} ET_{\bar{D}}$$

Donde la constante $t_{n_D-1, 0,025}$ se determina a partir de la distribución t de Student con

$$gl = n_D - 1$$

Los intervalos con otros grados de confianza (como el 90 %, 99 %, etc.) se construyen de forma análoga (utilizando $t_{0,05}$, $t_{0,005}$, etc.). El ejemplo siguiente ilustra el intervalo de confianza.

Ejemplo 8.2.2 Flujo sanguíneo

Para los datos de flujo sanguíneo, tenemos $gl = 8 - 1 = 7$. De la Tabla 4 obtenemos que $t_{7, 0,025} = 2,365$. Por tanto, el intervalo de confianza del 95 % para μ_D es

$$1,15 \pm (2,365) \left(\frac{0,63}{\sqrt{8}} \right)$$

o

$$1,15 \pm 0,53$$

o

$$(0,62, 1,68)$$

■

Podemos también realizar un test t . Para contrastar la hipótesis nula

$$H_0 : \mu_D = 0$$

utilizamos el estadístico de contraste

$$t_s = \frac{\bar{d} - 0}{ET_{\bar{D}}}$$

Los valores críticos se obtienen de la distribución t de Student (Tabla 4) con $gl = n_D - 1$. El ejemplo siguiente ilustra el test t .

Ejemplo 8.2.3 Flujo sanguíneo

Para los datos de flujo sanguíneo, formulemos la hipótesis nula y la alternativa no direccional:

H_0 : el flujo medio de sangre en el miocardio es el mismo en el punto de partida y después de tomar cafeína.

H_A : el flujo medio de sangre en el miocardio es diferente después de tomar cafeína y en el punto de partida. o, utilizando símbolos,

$$H_0: \mu_D = 0$$

$$H_A: \mu_D \neq 0$$

Contrastemos H_0 frente a H_A con un nivel de significación de $\alpha = 0,05$. El estadístico de contraste es

$$t_s = \frac{1,15 - 0}{0,63/\sqrt{8}} = 5,16$$

De la Tabla 4, $t_{7,0,005} = 3,499$ y $t_{7,0,0005} = 5,408$. Rechazamos H_0 y encontramos que hay evidencia suficiente ($0,001 < P < 0,01$) para concluir que el flujo medio de sangre en el miocardio disminuye después de tomar cafeína. (Utilizando un computador se tendría un P valor de $P = 0,0013$). (Nótese que, incluso aunque hay evidencia significativa a favor de una disminución en el FSM después de tomar la cafeína, no podemos concluir que la cafeína sea la causa de dicha disminución. Por ejemplo, podría ser que el flujo sanguíneo disminuyera debido al paso del tiempo). ■

Resultado de ignorar el emparejamiento

Supongamos que se realiza un estudio utilizando un diseño pareado, pero que el emparejamiento se ignora en el análisis de los datos. Ese análisis no es válido porque asume que las muestras son independientes cuando de hecho no lo son. El análisis incorrecto puede ser engañoso, como ilustra el ejemplo siguiente.

Ejemplo 8.2.4 Valoración del hambre

Durante un estudio sobre la pérdida de peso se suministró a siete sujetos o bien la sustancia activa *m*-clorofenil-piperazina (mCFP) durante dos semanas y después un placebo durante otras dos semanas, o bien se suministró un placebo las dos primeras semanas y después mCFP durante las siguientes dos semanas. Como parte del estudio, se preguntó a los sujetos que clasificaran lo hambrientos que se sentían al terminar cada periodo de dos semanas. Los datos sobre valoración del hambre se muestran en la Tabla 8.2.2.

Tabla 8.2.2 Datos de valoración del hambre para nueve mujeres			
Sujeto	Valoración del hambre		
	Sustancia (mCFP) y_1	Placebo y_2	Diferencia $d = y_1 - y_2$
1	79	78	1
2	48	54	-6
3	52	142	-90
4	15	25	-10
5	61	101	-40
6	107	99	8
7	77	94	-17
8	54	107	-53
9	5	64	-59
Media	55	85	-30
DT	32	34	33

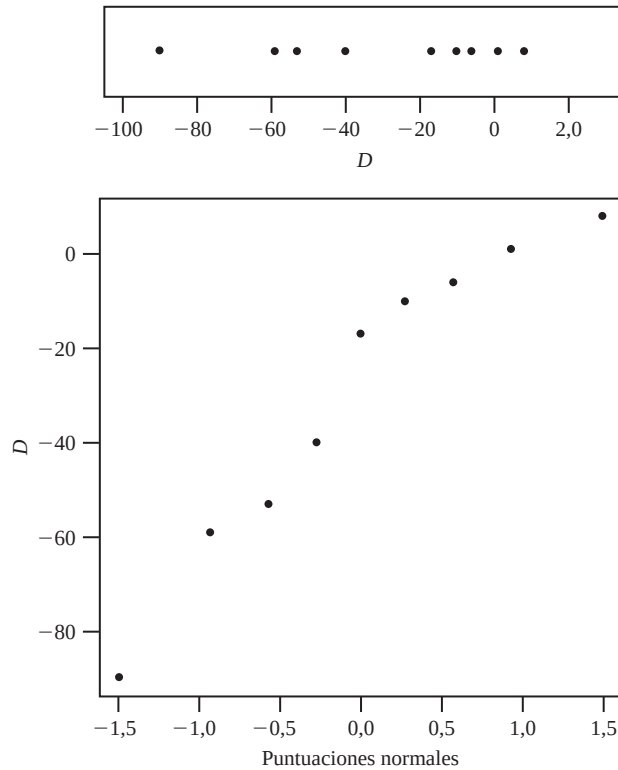


Figura 8.2.2 Diagrama de puntos de las diferencias en las valoraciones del hambre con la sustancia y con el placebo, junto con una gráfica de probabilidad normal de los datos

Para los datos de valoración del hambre, el ET de la diferencia media es

$$ET_{\bar{D}} = \frac{33}{\sqrt{9}} = 11$$

La Figura 8.2.2 muestra la distribución de las nueve diferencias muestrales.

Un contraste de

$$H_0: \mu_D = 0$$

frente a

$$H_A: \mu_D \neq 0$$

da un estadístico para el test de

$$t_s = \frac{-30 - 0}{11} = -2,72$$

Este estadístico del test tiene 8 grados de libertad. Utilizando un computador se obtiene el P valor de $P = 0,027$.

La Figura 8.2.3 muestra los datos de la sustancia y el placebo de forma separada. Hay considerable solapamiento entre las dos distribuciones. Este diagrama no muestra evidencia convincente de que la sustancia disminuye las valoraciones del hambre (tal como se determinan mediante el análisis pareado anterior) porque el diagrama no tiene en cuenta la naturaleza del emparejamiento de estos datos.

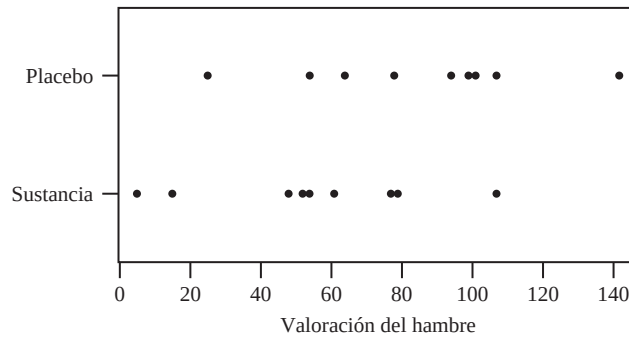


Figura 8.2.3 Diagramas de puntos en paralelo de las valoraciones del hambre con la sustancia y con el placebo

Observando separadamente los datos de la sustancia y del placebo, las DT de las dos muestras son $s_1 = 32$ y $s_2 = 34$. Si procedemos impropriamente como si las muestras fueran independientes y aplicamos la fórmula del ET del Capítulo 7, obtenemos

$$\begin{aligned} ET_{(\bar{y}_1 - \bar{y}_2)} &= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ &= \sqrt{\frac{32^2}{9} + \frac{34^2}{9}} = 15,6 \end{aligned}$$

Este ET es bastante más grande que el valor ($ET_{\bar{D}} = 11$) que calculamos utilizando el emparejamiento.

Continuando (erróneamente) como si las muestras fueran independientes, el estadístico de contraste es

$$t_s = \frac{55 - 85}{15,6} = -1,92$$

El P valor de este contraste es 0,075, que es mucho mayor que el P valor del contraste correcto, 0,027.

Para continuar comparando los análisis emparejado y no emparejado, consideremos el intervalo de confianza del 95 % para $(\mu_1 - \mu_2)$. En el análisis no emparejado, la fórmula (6.7.1) da un resultado de $15,9 \approx 16$ grados de libertad. Esto produce un multiplicador de $t_{16, 0,025} = 2,121$ y resulta en un intervalo de confianza de

$$(55 - 85) \pm (2,121)(15,6)$$

o

$$-30 \pm 33,1$$

o

$$(-63,1, 3,1)$$

Este intervalo de confianza erróneo es más amplio que el intervalo de confianza correcto obtenido mediante un análisis pareado. Un análisis emparejado produce el intervalo más estrecho

$$-30 \pm (2,306)(11)$$

o

$$-30 \pm 25,4$$

o

$$(-55,4, -4,6)$$

El intervalo para las muestras pareadas es más estrecho porque utiliza un ET menor. Este efecto está ligeramente compensado por un valor mayor de $t_{0,025}$ (2,306 frente a 2,121).

¿Por qué es el ET de las muestras pareadas menor que el ET de las muestras independientes calculadas con los mismos datos (ET = 11 frente a ET = 15,6)? La Tabla 8.2.2 revela la razón. Los datos muestran que hay una variación grande de un sujeto al siguiente. Por ejemplo, el sujeto 4 tiene valoraciones del hambre bajas (tanto con la sustancia como con el placebo) y el sujeto 6 tiene valores altos. La fórmula del ET para muestras independientes incorpora toda esta variación (expresada a través de s_1 y s_2). En el caso de las muestras pareadas, la variación de la valoración del hambre entre los sujetos no tiene influencia en los cálculos porque solo los utilizan los valores de D . Utilizando cada sujeto como su propio control, el investigador ha aumentado la precisión del experimento. Pero si el emparejamiento se ignora en el análisis, esta precisión extra se pierde. ■

El ejemplo anterior ilustra la ganancia de precisión que se puede obtener con un diseño pareado acoplado con un análisis pareado. La selección entre un diseño pareado y uno no pareado se comentará en la Sección 8.3.

Condiciones de validez del análisis basado en la t de Student

Las condiciones para la validez del test t y del intervalo de confianza para muestras pareadas son las siguientes:

1. Debe ser razonable considerar las *diferencias* (los valores de D) como una muestra aleatoria de una población grande.
2. La distribución poblacional de los valores de D debe ser normal. Los métodos son aproximadamente válidos si la distribución poblacional es aproximadamente normal o si el tamaño de la muestra (n_D) es grande.

Las condiciones anteriores son las mismas que las que se dieron en el Capítulo 6. En el presente caso, las condiciones se aplican a los valores de D porque el análisis se basa en dichos valores. La verificación de las condiciones se puede realizar como se describió en el Capítulo 6. Primero se debería comprobar el diseño para asegurar que los valores de D son independientes entre sí, y especialmente que no hay estructuras jerárquicas dentro de los valores de D . (Nótese, sin embargo, que los valores de Y_1 no son independientes de los valores de Y_2 debido al emparejamiento). Segundo, un histograma o un diagrama de puntos de los valores de D puede proporcionar una comprobación estimativa de la normalidad aproximada. También se puede utilizar una gráfica de probabilidad normal para evaluar la normalidad.

Nótese que no se requiere la normalidad de los Y_1 ni de los Y_2 , porque el análisis depende solo de los valores de D . El siguiente ejemplo muestra un caso en el cual los Y_1 e Y_2 no están distribuidas normalmente, pero sí lo están los valores de D .

Ejemplo 8.2.5 Ardillas

Si caminamos hacia una ardilla que está en el suelo, finalmente correrá al árbol más cercano por seguridad. Un investigador se preguntaba si se podría acercar a la ardilla más de lo cerca que la ardilla estaba del árbol más cercano antes de que la ardilla empezara a correr. Realizó 11 observaciones, que se muestran en la Tabla 8.2.3. La Figura 8.2.4 muestra que la distribución de las distancias desde la ardilla a la persona parece ser razonablemente normal, pero que las distancias desde la ardilla al árbol están lejos de estar normalmente distribuidas. Sin embargo, el panel (c) de la Figura 8.2.4 muestra que las 11 diferencias cumplen la condición de normalidad. Como el test t para muestras pareadas analiza las diferencias, sería válido aquí realizar dicho test t (o calcular un intervalo de confianza)³. ■

Resumen de fórmulas

Para disponer de una referencia conveniente, resumimos las fórmulas de los métodos basados en la t de Student para muestras pareadas.

Tabla 8.2.3 Distancias (en pulgadas) a la persona y al árbol cuando la ardilla empezó a correr			
Ardilla	A la persona y_1	Al árbol y_2	Diferencia $d = y_1 - y_2$
1	81	137	-56
2	178	34	144
3	202	51	151
4	325	50	275
5	238	54	184
6	134	236	-102
7	240	45	195
8	326	293	33
9	60	277	-217
10	119	83	36
11	189	41	148
Media	190	118	72
DT	89	101	148

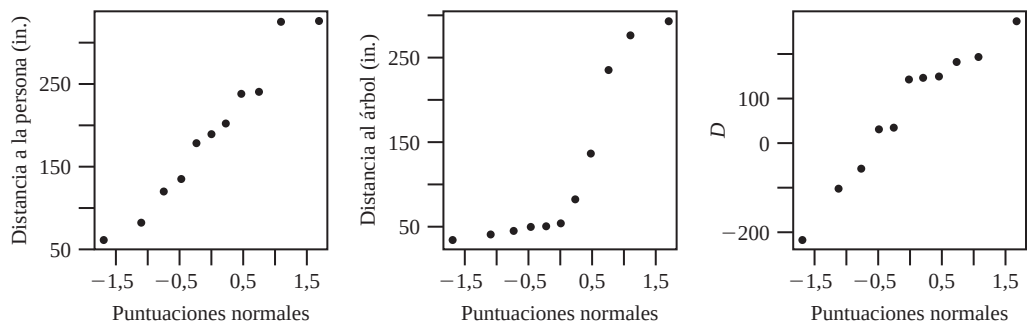


Figura 8.2.4 Gráficas de probabilidad normal de la distancia de las ardillas a la persona y de las ardillas al árbol

Error típico de $\bar{D}$

$$ET_{\bar{D}} = \frac{s_D}{\sqrt{n_D}}$$

Test t

$$H_0: \mu_D = 0$$

$$t_s = \frac{\bar{d} - 0}{ET_{\bar{D}}}$$

Intervalo de confianza del 95 % para μ_d

$$\bar{d} \pm t_{0,025} ET_{\bar{D}}$$

Los intervalos con otros niveles de confianza (por ejemplo, del 90 % o del 99 %) se construyen análogamente (es decir, utilizando $t_{0,05}$ o $t_{0,005}$).

Ejercicios 8.2.1-8.2.11

8.2.1 En un experimento agrónomo de campo se subdividieron bloques de tierra en dos parcelas de 346 pies cuadrados cada una. Cada bloque proporcionó dos observaciones emparejadas: una para cada una de las variedades de trigo. Las cosechas de trigo de cada parcela (lb) se muestran en la tabla que se presenta a continuación⁴.

- (a) Calcule el error típico de la diferencia de medias entre las variedades.
- (b) Realice un contraste para la diferencia de las variedades utilizando un test t para muestras emparejadas con $\alpha = 0,05$. Utilice una alternativa no direccional.
- (c) Realice un contraste para la diferencia de las variedades de forma incorrecta, utilizando un contraste para muestras independientes. Compare con el resultado del apartado (b).

	Variedad		
Bloque	1	2	Diferencia
1	32,1	34,5	-2,4
2	30,6	32,6	-2,0
3	33,7	34,6	-0,9
4	29,7	31,0	-1,3
Media	31,52	33,17	-1,65
DT	1,76	1,72	0,68

8.2.2 En un experimento para comparar dos dietas de engorde de ganado vacuno, se eligieron nueve parejas de animales de la manada. Los animales se ajustaron lo más posible con respecto a factores hereditarios. Los miembros de cada pareja se asignaron aleatoriamente a cada una de las dos dietas. La tabla siguiente muestra las ganancias de peso (lb) de los animales durante un periodo de prueba de 140 días con la dieta uno (Y_1) y con la dieta dos (Y_2)⁵.

Pareja	Dieta 1	Dieta 2	Diferencia
1	596	498	98
2	422	460	-38
3	524	468	56
4	454	458	-4
5	538	530	8
6	552	482	70
7	478	528	-50
8	564	598	-34
9	556	456	100
Media	520,4	497,6	22,9
DT	57,1	47,3	59,3

- (a) Calcule el error típico de la diferencia de medias.

- (b) Contraste una diferencia entre las dietas utilizando un test t para muestras pareadas con $\alpha = 0,10$. Utilice una alternativa no direccional.
- (c) Construya un intervalo de confianza del 90 % para μ_D .
- (d) Interprete el intervalo de confianza del apartado (c) en el contexto de este ejercicio.

8.2.3 El monofosfato adenosina cíclico (cAMP) es una sustancia mediadora de la respuesta celular a las hormonas. En un estudio de maduración de óvulos de la rana *Xenopus laevis* se dividieron ovocitos de cuatro hembras en dos grupos. Un grupo se expuso a progesterona y el otro no. Tras dos minutos, se midió el contenido de cAMP de cada grupo, con los resultados que se muestran en la tabla siguiente⁶. Utilice un test t para investigar el efecto de la progesterona sobre el cAMP. Considere H_A no direccional y $\alpha = 0,10$.

	cAMP (pmol/ovocito)		
Rana	Control	Progesterona	d
1	6,01	5,23	0,78
2	2,28	1,21	1,07
3	1,51	1,40	0,11
4	2,12	1,38	0,74
Media	2,98	2,31	0,68
DT	2,05	1,95	0,40

8.2.4 La tabla siguiente muestra la cantidad de pérdida de peso (kg) para los nueve sujetos del Ejemplo 8.2.4 cuando tomaron la sustancia mCFP y cuando tomaron un placebo². (Nótese que si un sujeto ganó peso, entonces la pérdida de peso que se presenta es negativa, como en el caso del sujeto 2, que ganó 0,3 kg con el placebo). Utilice un test t para investigar la afirmación de que la sustancia mCFP afecta a la pérdida de peso. Considere H_A no direccional y $\alpha = 0,01$.

	Cambio de peso		
Sujeto	mCFP	Placebo	Diferencia
1	1,1	0,0	1,1
2	1,3	-0,3	1,6
3	1,0	0,6	0,4
4	1,7	0,3	1,4
5	1,4	-0,7	2,1
6	0,1	-0,2	0,3
7	0,5	0,6	-0,1
8	1,6	0,9	0,7
9	-0,5	-2,0	1,5
Media	0,91	-0,09	1,00
DT	0,74	0,88	0,72

8.2.5 En referencia al Ejercicio 8.2.4:

- (a) Construya un intervalo de confianza al 99 % para μ_D .
 (b) Interprete el intervalo de confianza del apartado (a) en el contexto de este ejercicio.

8.2.6 Bajo ciertas condiciones, la estimulación eléctrica de una res de ganado hará que la carne sea más tierna. En un estudio sobre este efecto, se dividió por la mitad un conjunto de reses de ganado sacrificadas para carne. Una mitad se sometió a una breve corriente eléctrica y la otra mitad se dejó como control sin tratar. Se tomó un filete de cada una de las mitades y se probó lo tierno que era de varias formas. En una de las pruebas, el investigador tuvo un espécimen de tejido conectivo (colágeno) de filete y determinó la temperatura a la cual el tejido se reducía. Una pieza tierna de carne tiende a tener una temperatura de reducción del colágeno más baja. Los datos se muestran en la tabla siguiente⁷.

- (a) Construya un intervalo de confianza del 95 % para la diferencia de medias entre la mitad tratada y la mitad de control.
 (b) Construya un intervalo de confianza del 95 % de forma errónea, utilizando el método de muestras independientes. ¿En qué se diferencia este intervalo del obtenido en el apartado (a)?

Temperatura de reducción del colágeno (°C)			
Res	Mitad tratada	Mitad de control	Diferencia
1	69,50	70,00	-0,50
2	67,00	69,00	-2,00
3	70,75	69,50	1,25
4	68,50	69,25	-0,75
5	66,75	67,75	-1,00
6	68,50	66,50	2,00
7	69,50	68,75	0,75
8	69,00	70,00	-1,00
9	66,75	66,75	0,00
10	69,00	68,50	0,50
11	69,50	69,00	0,50
12	69,00	69,75	-0,75
13	70,50	70,25	0,25
14	68,00	66,25	1,75
15	69,00	68,25	0,75
Media	68,750	68,633	0,117
DT	1,217	1,302	1,118

8.2.7 En referencia al Ejercicio 8.2.6, utilice un test t para contrastar la hipótesis nula de que no hay efecto contra la

hipótesis alternativa de que el tratamiento eléctrico tiende a reducir la temperatura de reducción del colágeno. Use $\alpha = 0,10$.

8.2.8 La tricotilomanía es una enfermedad psiquiátrica que hace que sus víctimas tengan una tendencia irresistible de tirar de su propio cabello. En un estudio con 13 mujeres, se compararon dos medicamentos como tratamientos de la tricotilomanía. Cada mujer tomó clomipramina durante un periodo de tiempo y desipramina durante otro periodo de tiempo en un experimento doble ciego. Durante cada periodo de tiempo se midieron las puntuaciones de cada mujer en una escala de afectación de la tricotilomanía, indicando una puntuación mayor una afectación mayor⁸. El promedio de las 13 medidas para la clomipramina fue de 6,2. El promedio de las 13 medidas para desipramina fue de 4,2. Un test t de muestras pareadas dio un valor de $t_s = 2,47$ y un P valor de dos colas de 0,03. Interprete el resultado del test t . Es decir, ¿qué indica el texto sobre la clomipramina, la desipramina y la tricotilomanía?

8.2.9 Un científico realizó un estudio sobre la frecuencia de canto de su periquito. Apuntó el número de trinos distintos que hizo el periquito en un periodo de 30 minutos, algunas veces cuando la habitación estaba en silencio y algunas veces cuando se había puesto música. Los datos se muestran en la tabla siguiente⁹. Construya un intervalo de confianza del 95 % para el número medio de trinos (cada 30 minutos) cuando se está poniendo música con respecto a cuando no se está poniendo música.

Trinos en 30 minutos			
Día	Con música	Sin música	Diferencia
1	12	3	9
2	14	1	13
3	11	2	9
4	13	1	12
5	20	5	15
6	14	3	11
7	10	0	10
8	12	2	10
9	8	6	2
10	13	3	10
11	14	2	12
12	15	4	11
13	12	3	9
14	13	2	11
15	8	0	8

(Continúa)

Día	Trinos en 30 minutos		
	Con música	Sin música	Diferencia
16	18	5	13
17	15	3	12
18	12	2	10
19	17	2	15
20	15	4	11
21	11	3	8
22	22	4	18
23	14	2	12
24	18	4	14
25	15	5	10
26	8	1	7
27	13	2	11
28	16	3	13
Media	13,7	2,8	10,9
DT	3,4	1,5	3,0

8.2.10 Considere datos del Ejercicio 8.2.9. Hay dos *outliers* en las 28 diferencias: el valor más pequeño, que es 2, y el valor más grande, que es 18. Elimine estas dos observaciones y construya un intervalo de confianza del 95 % para el incremento de las medias, utilizando las 26 observaciones restantes. ¿Tienen mucho efecto los *outliers* en el intervalo de confianza?

8.2.11 Invente un conjunto de datos emparejados, consistente en cinco parejas de observaciones para las que $\bar{y}_1$ e $\bar{y}_2$ no sean iguales, y $ET_{\bar{y}_1} > 0$ y $ET_{\bar{y}_2} > 0$, pero $ET_{\bar{D}} = 0$.

8.3 El diseño pareado

Idealmente, en un diseño pareado los miembros de una pareja son relativamente similares entre sí (es decir, más similares entre sí que a los miembros de otras parejas) con respecto a variables externas. La ventaja de esta disposición es que, cuando se comparan los miembros de una pareja, la comparación está libre de variaciones externas que originan diferencias entre parejas. Ampliaremos este tema después de dar algunos ejemplos.

Ejemplos de diseños pareados

Los diseños pareados pueden surgir de diferentes formas, entre las que se encuentran las siguientes:

Experimentos en los que las unidades experimentales similares forman parejas.

Estudios observacionales de gemelos idénticos.

Medidas repetidas del mismo individuo en dos instantes diferentes.

Emparejamiento por tiempo.

Experimentos con parejas de unidades Muchas veces los investigadores que desean comparar dos tratamientos forman primero parejas de unidades experimentales (parejas de animales, parejas de parcelas de tierra, etc.) que son similares (por ejemplo, animales de la misma edad y sexo o parcelas de tierra con el mismo tipo de suelo y exposición al viento, a la lluvia y al sol). Seguidamente, un miembro de la pareja se asigna aleatoriamente para recibir el primer tratamiento y el otro miembro se asigna al segundo tratamiento. Se presenta a continuación un ejemplo.

Ejemplo 8.3.1 Fertilizantes para berenjenas

En un experimento de invernadero para comparar dos tratamientos de fertilizantes para berenjenas, un conjunto de plantas en tiesto de dicho invernadero se agrupan en parejas, de forma que las dos plantas de la misma pareja estén expuestas a la misma cantidad de luz solar, la misma temperatura, etc. Dentro de cada pareja, una planta (elegida aleatoriamente) recibirá el tratamiento 1 y la otra recibirá el tratamiento 2. ■

Estudios observacionales Como se indicó en la Sección 7.4, son preferibles los experimentos aleatorizados a los estudios observacionales, debido a las muchas variables de confusión que pueden aparecer en un estudio observacional. Un estudio observacional nos puede decir que X e Y están *asociadas*, pero solo un experimento apropiado nos puede decir si X es *causa* de Y . Si no es posible realizar un experimento y se debe realizar un estudio observacional, entonces es preferible (aunque raramente posible) utilizar gemelos idénticos como unidades observacionales. Por ejemplo, en un estudio del efecto de «fumador pasivo» sería ideal incorporar varios conjuntos de gemelos que no fumaran de forma que, en cada pareja, uno de los gemelos viviera con un fumador y el otro con un no fumador. Debido a que los conjuntos de gemelos idénticos están raramente, si es que alguna vez lo están, disponibles, se utilizan frecuentemente los diseños de **parejas igualadas**, en los que los dos grupos se igualan con respecto a varias variables externas¹⁰. He aquí un ejemplo.

Ejemplo 8.3.2 Fumar y cáncer de pulmón

En un estudio de control de casos de cáncer de pulmón se identificaron 100 pacientes de cáncer de pulmón. En cada caso se escogió un control que estaba igualado individualmente al caso con respecto a la edad, el sexo y el nivel de educación. Se compararon los hábitos de fumar de los casos y de los controles. ■

Medidas repetidas Muchas investigaciones biológicas utilizan medidas repetidas realizadas sobre el mismo individuo en momentos diferentes. Entre ellas se encuentran estudios sobre crecimiento y desarrollo, estudios sobre procesos biológicos y estudios en los que las medidas se hacen antes y después de la aplicación de un cierto tratamiento. Cuando solo se utilizan dos momentos, las medidas están emparejadas, como en el Ejemplo 8.1.1. Se presenta a continuación otro ejemplo.

Ejemplo 8.3.3 Ejercicio y triglicéridos en suero

Los triglicéridos son componentes de la sangre que se piensa tienen un papel en la enfermedad de la arteria coronaria. Para ver si el ejercicio regular podría reducir los niveles de triglicéridos, unos investigadores midieron la concentración de triglicéridos en el suero sanguíneo de siete voluntarios varones, antes y después de la participación en un programa de ejercicio de 10 semanas. Los resultados se muestran en la Tabla 8.3.1¹¹. Nótese que hay una variación considerable de un participante a otro. Por ejemplo, el participante 1 tiene niveles de triglicéridos relativamente bajos tanto antes como después, mientras que el participante 3 tiene niveles relativamente altos. ■

Tabla 8.3.1 Triglicéridos en suero (mmol/L)		
Participante	Antes	Después
1	0,87	0,57
2	1,13	1,03
3	3,14	1,47
4	2,14	1,43
5	2,98	1,20
6	1,18	1,09
7	1,60	1,51

Emparejamiento por tiempo En algunas situaciones, las parejas se forman implícitamente cuando se repiten medidas que se hacen en instantes diferentes. Sigue a continuación un ejemplo.

Ejemplo 8.3.4 Crecimiento de virus

En una serie de experimentos con cierto virus (mengovirus), un microbiólogo midió el crecimiento de dos cepas de virus (una cepa mutante y una cepa no mutante) sobre células de ratón en placas de Petri. Se realizaron experimentos repetidos en 19 días diferentes. Los datos se muestran en la Tabla 8.3.2. Cada número representa el crecimiento total de los virus en una sola placa en 24 horas¹².

Tabla 8.3.2 Crecimiento de virus en 24 horas

Replicación	Cepa no mutante	Cepa mutante	Replicación	Cepa no mutante	Cepa mutante
1	160	97	11	61	15
2	36	55	12	14	10
3	82	31	13	140	150
4	100	95	14	68	44
5	140	80	15	110	31
6	73	110	16	37	14
7	110	100	17	95	57
8	180	100	18	64	70
9	62	6	19	58	45
10	43	7			

Nótese que hay una considerable variación de una replicación a otra. Por ejemplo, la replicación 1 tuvo valores relativamente grandes (160 y 97), mientras que la replicación 2 tuvo valores relativamente pequeños (36 y 55). Esta variación entre réplicas surge de pequeñas variaciones inevitables en las condiciones experimentales. Por ejemplo, tanto el crecimiento de los virus como la técnica de medida son altamente sensibles a las condiciones ambientales como la temperatura y la concentración de CO_2 de la incubadora. No se pueden evitar ligeras fluctuaciones de las condiciones experimentales, y estas fluctuaciones causan la variación que reflejan los datos. En esta clase de situación la ventaja de utilizar las dos cepas concurrentemente (es decir, en parejas) es particularmente destacada. ■

Los Ejemplos 8.3.3 y 8.3.4 involucran ambas medidas en instantes diferentes. Pero nótese que la estructura del emparejamiento en los dos ejemplos es completamente diferente. En el Ejemplo 8.3.3 los miembros de una pareja son medidas en el mismo individuo en dos instantes diferentes, mientras que en el Ejemplo 8.3.4 los miembros de una pareja son medidas en dos placas de Petri en el mismo instante. No obstante, en ambos ejemplos el principio del emparejamiento es el mismo: los miembros de una pareja son similares entre sí con respecto a variables externas. En el Ejemplo 8.3.4 el tiempo es una variable externa, mientras que en el Ejemplo 8.3.3 la comparación entre dos instantes (antes y después) es de interés primordial y la variación entre sujetos es externa.

Objetivos del emparejamiento

El emparejamiento en un diseño experimental puede servir para reducir el sesgo, para aumentar la precisión o ambas cosas. Generalmente el propósito principal del emparejamiento es aumentar la precisión.

En la Sección 7.4 indicamos que emparejamiento o igualación puede reducir el sesgo controlando la variación debida a variables externas. Las variables utilizadas en la igualación se equilibran necesariamente en los dos grupos que se van a comparar y, por tanto, no pueden distorsionar la comparación. Por ejemplo, si se comparan dos grupos de parejas de personas igualadas por edad, entonces cualquier comparación entre los dos grupos estará libre de sesgo debido a una diferencia en la distribución de las edades.

En experimentos aleatorizados, donde se puede controlar el sesgo mediante la asignación aleatoria, una razón importante para el emparejamiento es aumentar la precisión. El emparejamiento efectivo incrementa la precisión porque aumenta la información disponible en el experimento. Un análisis apropiado, que extraiga esta información extra, conducirá a contrastes con más potencia e intervalos de confianza más estrechos. Por tanto, un experimento pareado de forma efectiva es más eficiente. Produce más información que un experimento no emparejado con el mismo número de observaciones.

Vimos un ejemplo de emparejamiento efectivo en los datos de valoración del hambre del Ejemplo 8.2.4. El emparejamiento era efectivo porque una buena parte de la variación de las medidas era debida a la variación entre los sujetos, que no influye en la comparación entre los tratamientos. Como resultado, el experimento produjo información más precisa

sobre las diferencias de tratamientos que la que hubiera producido un experimento no emparejado comparable (es decir, un experimento que comparara las valoraciones del hambre de nueve mujeres que habían recibido mCFP con las valoraciones del hambre de nueve mujeres de control diferentes que habían recibido el placebo).

La efectividad de un emparejamiento dado se puede representar visualmente mediante un diagrama de dispersión de Y_1 frente a Y_2 . Cada punto del diagrama de dispersión representa una pareja (Y_1, Y_2) . La Figura 8.3.1 muestra un diagrama de dispersión de los datos de crecimiento de virus del Ejemplo 8.3.4, junto con un diagrama de caja de las diferencias. Cada punto del diagrama de dispersión representa una replicación. Nótese que los puntos del diagrama de dispersión muestran una tendencia creciente definida. Esta tendencia creciente indica la efectividad del emparejamiento. Las medidas de la misma replicación (es decir, del mismo día) tienen más en común que las medidas de replications diferentes, de forma que una réplica con un valor relativamente alto de Y_1 tiende a tener un valor relativamente alto de Y_2 , y lo mismo para los valores bajos.

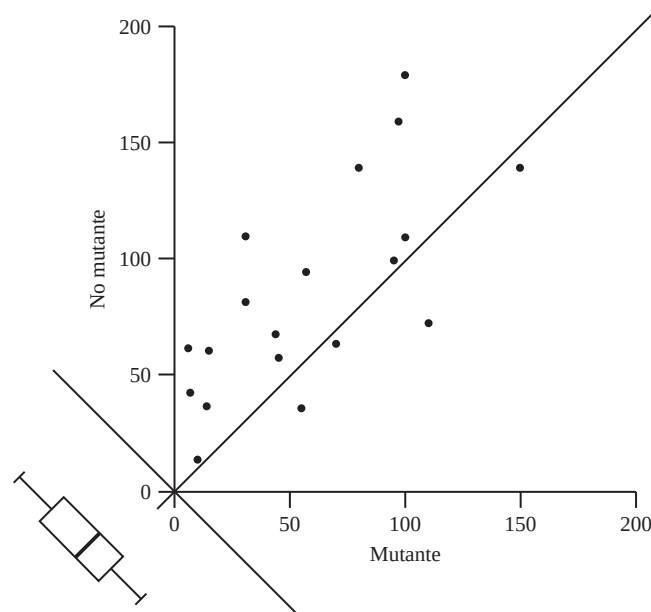


Figura 8.3.1 Diagramas de dispersión de los datos de crecimiento de virus, con un diagrama de caja de las diferencias

Nótese que el emparejamiento es una estrategia de *diseño*, no de análisis y, por tanto, se realiza *antes* de observar los valores de Y . No es correcto utilizar las propias observaciones para formar parejas. Una manipulación de los datos como esa podría distorsionar severamente los resultados experimentales y se podría considerar un fraude científico.

Diseño en parejas aleatorizadas frente a diseño completamente aleatorizado

Al planificar un experimento aleatorizado, el investigador puede necesitar decidir entre un diseño pareado y un diseño que utilice asignación aleatoria sin emparejamiento, que se denomina diseño completamente aleatorio. Hemos dicho que el emparejamiento efectivo puede aumentar grandemente la precisión de un experimento. Por otra parte, el emparejamiento puede no ser efectivo en un experimento determinado, si la variable observada Y no está relacionada con los factores utilizados en dicho emparejamiento. Por ejemplo, supongamos que las parejas se igualan únicamente por edad, pero de hecho Y resulta no estar relacionada con la edad. Se puede demostrar que el emparejamiento inefectivo puede obtener realmente menos precisión que no emparejar en absoluto. Por ejemplo, en relación con un test t , el emparejamiento inefectivo no tendería a reducir el ET, sino que reduciría los grados de libertad, y el resultado neto sería una pérdida de potencia.

La decisión de utilizar un diseño pareado depende de consideraciones prácticas (el emparejamiento puede ser caro o difícil de manejar) y de consideraciones de precisión. Con respecto a la precisión, la decisión depende de cómo se espera que sea el emparejamiento de efectivo. El siguiente ejemplo ilustra este aspecto.

Ejemplo 8.3.5 Fertilizantes para berenjenas

Un horticultor está planificando un experimento de invernadero con plantas de berenjena sembradas en tiestos individuales. Se van a comparar dos tratamientos de fertilizante, la variable observada es Y = cosecha de berenjenas (libras). El investigador sabe que Y está influida por factores como la luz y la temperatura, que varían de un sitio a otro en los bancos del invernadero. La situación de los tiestos en posiciones del banco se podría realizar de acuerdo con un diseño completamente aleatorizado o de acuerdo con un diseño emparejado, como en el Ejemplo 8.3.1. Al decidir entre estas opciones, el investigador debe utilizar su conocimiento sobre lo efectivo que sería el emparejamiento (es decir, si los tiestos situados adyacentes en el banco producirían cosechas muy similares que los tiestos situados lejos). Si juzga que el emparejamiento no sería muy efectivo, puede optar por un diseño completamente aleatorizado. ■

Nótese que el emparejamiento efectivo *no* es lo mismo que simplemente mantener constantes las condiciones experimentales. El emparejamiento es una forma de *organizar* la variación inevitable que queda todavía, después que las condiciones experimentales se han hecho tan constantes como es posible. El emparejamiento ideal organiza la variación de forma que dicha variación dentro de cada pareja es mínima y entre parejas es máxima.

Selección del análisis

El análisis de datos se debe ajustar al diseño del estudio. Si el diseño es pareado, se debe utilizar un análisis para muestras pareadas. Si el diseño es no emparejado, se debería utilizar un análisis para muestras independientes (como en el Capítulo 7).

Nótese que la información adicional que está disponible por la realización de un diseño pareado efectivo se *pierde completamente* si se utiliza un análisis no emparejado. (Vimos una ilustración de esto en el Ejemplo 8.2.4). Por tanto, el diseño pareado no aumenta la eficiencia, a menos que esté acompañado por un análisis de muestras pareadas.

Ejercicios 8.3.1-8.3.4

8.3.1 (Ejercicio de muestreo) Este ejercicio ilustra la aplicación del diseño pareado a la población de 100 elipses (que se muestra en el Ejercicio 3.1.1). La tabla que se presenta a continuación muestra la agrupación de las 100 elipses en 50 parejas.

Pareja	Números de ID de las elipses			Pareja	Números de ID de las elipses			Pareja	Números de ID de las elipses		
01	20	45		18	11	46		35	16	66	
02	03	49		19	09	29		36	18	58	
03	07	27		20	19	39		37	30	50	
04	42	82		21	00	10		38	76	86	
05	81	91		22	40	55		39	17	83	
06	38	72		23	21	56		40	04	52	
07	60	70		24	08	62		41	12	64	
08	31	61		25	24	78		42	23	57	

(Continúa)

Pareja	Números de ID de las elipses			Pareja	Números de ID de las elipses			Pareja	Números de ID de las elipses		
09	77	89		26	67	93		43	98	99	
10	01	41		27	35	80		44	36	96	
11	14	48		28	74	88		45	44	84	
12	59	87		29	94	97		46	06	51	
13	22	68		30	02	28		47	85	90	
14	47	79		31	26	71		48	37	63	
15	05	95		32	25	65		49	43	69	
16	53	73		33	15	75		50	34	54	
17	13	33		34	32	92					

Para apreciar mejor este ejercicio, imagine la siguiente construcción experimental. Se desea investigar el efecto de cierto tratamiento, T , sobre el organismo *C. ellipticus*. Se observará la variable Y = longitud. Se puede medir cada individuo solo una vez y, por tanto, comparar n tratamientos

individuales con n controles sin tratamiento. Se sabe que los individuos disponibles para el experimento son de varias edades, y se sabe que la edad está relacionada con la longitud, de forma que se han formado 50 parejas igualadas por edad, algunas de las cuales serán utilizadas en el experimento. El objetivo del emparejamiento es aumentar la potencia del experimento eliminando la variación aleatoria debida a la edad. (Por supuesto, las elipses no tienen realmente edad, pero el emparejamiento que se muestra en la tabla se ha construido de forma que *simula* la igualación por edad).

- Utilice dígitos aleatorios (de la Tabla 1 o de su calculadora) para elegir una muestra aleatoria de cinco parejas de elipses.
- Para cada pareja, utilice dígitos aleatorios (o lance una moneda) para asignar aleatoriamente un miembro al tratamiento (T) y otro al control (C).
- Mida las longitudes de las 10 elipses. Seguidamente, para simular un efecto del tratamiento, añada 6 mm a cada longitud del grupo T.
- Aplique a los datos un test t para muestras pareadas. Utilice una alternativa no direccional y $\alpha = 0,05$.
- ¿Produjo el análisis del apartado (d) un error de Tipo II?

8.3.2 (Continuación del Ejercicio 8.3.1) Aplique un test t de muestras independientes a sus datos. Utilice una alternativa no direccional y $\alpha = 0,05$. ¿Produce este análisis un error de Tipo II?

8.3.3 (Ejercicio de muestreo) En referencia al Ejercicio 8.3.1, imagine que no es práctico realizar un experimento de muestras pareadas (quizá porque no se pueden medir las edades de los individuos), por lo que se decide utilizar un experimento completamente aleatorizado para evaluar el tratamiento T.

- Utilice dígitos aleatorios (de la Tabla 1 o de su calculadora) para escoger una muestra aleatoria de 10 individuos de la población de elipses (que se muestra en el Ejercicio 3.1.1). De las 10, asigne aleatoriamente 5 a T y 5 a C. (O, de forma equivalente, seleccione aleatoriamente 5 elementos de la población para que reciban T y 5 para que reciban C).
- Mida las longitudes de las 10 elipses. Seguidamente, simule un efecto del tratamiento, añadiendo 6 mm a cada longitud del grupo T.
- Aplique un test de muestras independientes a los datos. Utilice una alternativa no direccional y $\alpha = 0,05$.
- ¿Produjo el análisis del apartado (c) un error de Tipo II?

8.2.4 En referencia a cada uno de los siguientes ejercicios, construya un diagrama de dispersión de los datos. ¿Indica la apariencia del diagrama de dispersión que el emparejamiento fue efectivo?

- Ejercicio 8.2.1.
- Ejercicio 8.2.2.
- Ejercicio 8.2.6.

8.4 El test de signos

El **test de signos** es un test no paramétrico que se puede utilizar para comparar dos muestras pareadas. No es particularmente potente, pero es muy flexible en su aplicación y es especialmente simple de utilizar y de entender.

Método

Como el test t de muestras pareadas, el test de signos se basa en las diferencias

$$D = Y_1 - Y_2$$

La única información que se utiliza en el test de signos es el signo (positivo o negativo) de cada diferencia. Si las diferencias son preponderantemente de un signo, esto se toma como evidencia a favor de la hipótesis alternativa. Los ejemplos siguientes ilustran el test de signos.

Ejemplo 8.4.1 Injertos de piel

La piel de los cadáveres puede servir para proporcionar injertos de piel temporales a pacientes con quemaduras graves. Cuanto más tiempo sobrevive un injerto antes de su inevitable rechazo por el sistema inmune, más beneficios para el paciente. Un equipo médico investigó la utilidad de ajustar el injerto al paciente con respecto al sistema antigénico A-LH (Antígeno Leucocitario Humano). Cada paciente recibió dos injertos, uno con alta compatibilidad A-LH y otro con baja compatibilidad. Los tiempos de supervivencia (en días) de los injertos de piel se muestran en la Tabla 8.4.1¹³.

Nótese que no se puede aplicar aquí un test t porque dos de las observaciones son incompletas. El paciente 3 murió con un injerto todavía vivo y la observación del paciente 10 fue incompleta por una razón no especificada. No obstante, podemos realizar un test de signos, ya que el test de signos depende solamente del signo de la diferencia para cada paciente y sabemos que $Y_1 - Y_2$ es positiva para ambos pacientes.

Tabla 8.4.1 Tiempos de supervivencia de injertos de piel			
Paciente	Compatibilidad A-LH		
	Alta y_1	Baja y_2	Signo de $d = y_1 - y_2$
1	37	29	+
2	19	13	+
3	57 +	15	+
4	93	26	+
5	16	11	+
6	23	18	+
7	20	26	-
8	63	43	+
9	29	18	+
10	60 +	42	+
11	18	19	-

Realicemos un test de signos para comparar los tiempos de supervivencia de los dos conjuntos de injertos de piel utilizando $\alpha = 0,05$. Una hipótesis de investigación (alternativa) direccional es adecuada en este experimento:

H_A : los injertos de piel tienden a durar más cuando la compatibilidad A-LH es grande.

La hipótesis nula es

H_0 : la distribución del tiempo de supervivencia es la misma para compatibilidad alta que para compatibilidad baja.

El primer paso es determinar los siguientes valores:

N_+ = Número de diferencias positivas

N_- = Número de diferencias negativas

Como H_A es direccional y predice que la mayoría de las diferencias serán positivas, el estadístico de contrastes B_s es

$$B_s = N_+$$

Para los datos presentes, tenemos

$$N_+ = 9$$

$$N_- = 2$$

$$B_s = 9$$

El siguiente paso es obtener el P valor. Utilizamos la letra B para denominar al estadístico de contraste B_s porque la distribución de B_s está basada en la distribución binomial. Sea p la probabilidad de que una diferencia sea positiva. Si la hipótesis nula es cierta, entonces $p = 0,5$. Por tanto, la distribución nula de B_s es binomial con $n = 11$ y $p = 0,5$. Es decir, la hipótesis nula implica que el signo de cada diferencia es como el resultado de lanzar una moneda, donde las caras corresponderían a una diferencia positiva y las cruces a una diferencia negativa.

Para los datos de injerto de piel, el P valor del contraste es la probabilidad de obtener 9 o más diferencias positivas en 11 pacientes si $p = 0,5$. Esto es igual a la probabilidad de que una variable aleatoria binomial con $n = 11$ y $p = 0,5$ sea mayor o igual que 9. Utilizando la fórmula binomial del Capítulo 3, o un computador, obtenemos que esta probabilidad es 0,03272*.

* Posteriormente en esta sección veremos cómo utilizar una tabla para calcular estos valores. Sin embargo, se ha seguido la sección opcional sobre la distribución binomial, se puede calcular esta probabilidad utilizando la fórmula binomial

$${}_{11}C_9(0,5)^9(0,5)^2 + {}_{11}C_{10}(0,5)^{10}(0,5)^1 + {}_{11}C_{11}(0,5)^{11} = 0,02686 + 0,00537 + 0,00049 = 0,03272$$

Como el P valor es menor que α , obtenemos evidencia significativa en el sentido de que los injertos de piel tienden a durar más cuando la compatibilidad A-LH es alta que cuando es baja. ■

Ejemplo 8.4.2 Crecimiento de virus

La Tabla 8.4.2 muestra los datos de crecimiento de virus del Ejemplo 8.3.4, junto con el signo que las diferencias.

Tabla 8.4.2 Crecimiento de virus tras 24 horas							
Replicación	Cepa no mutante y_1	Cepa mutante y_2	Signo de $d = y_1 - y_2$	Replicación	Cepa no mutante y_1	Cepa mutante y_2	Signo de $d = y_1 - y_2$
1	160	97	+	11	61	15	+
2	36	55	−	12	14	10	+
3	82	31	+	13	140	150	−
4	100	95	+	14	68	44	+
5	140	80	+	15	110	31	+
6	73	110	−	16	37	14	+
7	110	100	+	17	95	57	+
8	180	100	+	18	64	70	−
9	62	6	+	19	58	45	+
10	43	7	+				

Realicemos un contraste de signos para comparar el crecimiento de las dos cepas, utilizando $\alpha = 0,10$. La hipótesis nula y la alternativa no direccional son:

H_0 : las dos cepas del virus crecen igualmente bien.

H_A : una de las cepas crece mejor que la otra.

Para estos datos,

$$N_+ = 15$$

$$N_- = 4$$

Cuando la alternativa es no direccional, B_s se define como

$$B_s = \text{Máximo de } N_+ \text{ y } N_-$$

por lo que para los datos de crecimiento de virus,

$$B_s = 15$$

El P valor de este contraste es la probabilidad de obtener 15 o más éxitos, más la probabilidad de obtener cuatro o menos éxitos, en un experimento binomial con $n = 19$. Podríamos utilizar la fórmula binomial para calcular el P valor. Como alternativa, la Tabla 7 (al final del libro) contiene los valores críticos y los P valores para el test de signos. Utilizando la Tabla 7 con $n_D = 19$, obtenemos los valores críticos y los correspondientes P valores que se muestran en la Tabla 8.4.3:

Tabla 8.4.3 Valores críticos y P valores del test de signos cuando $n_D = 19$								
n_D	0,20	0,10	0,05	0,02	0,01	0,002	0,001	
19	13 0,167	14 0,064	15 0,019	15 0,019	16 0,004	17 0,0007	17 0,0007	

Utilizando la tabla podemos ver que para $B_s = 15$ el P valor es 0,019, por lo que existe evidencia significativa a favor de H_A . Es decir, rechazamos H_0 y concluimos que los datos proporcionan evidencia significativa de que la cepa no mutante crece mejor (tras 24 horas) que la cepa mutante del virus. ■

Acotación del P De la misma forma que el test de Wilcoxon-Mann-Whitney, el test de signos tiene una distribución nula discreta. Ciertas entradas de valores críticos de la Tabla 7 están en blanco, dado que en algunos casos los datos más extremos posibles no conducen a un P valor pequeño. La Tabla 7 tiene otra peculiaridad que no se comparte con el test Wilcoxon-Mann-Whitney: algunos valores críticos aparecen más de una vez en la misma fila debido a la discreción de la distribución nula.

Alternativa direccional Para utilizar la Tabla 7 si la hipótesis alternativa es direccional, procederemos con el procedimiento familiar en dos pasos:

Paso 1. Comprobar la direccionalidad (ver si los datos se desvían con respecto a H_0 en la dirección especificada por H_A).

- (a) Si no es así, el P valor es mayor que 0,50.
- (b) Si es así, continuar con el paso 2.

Paso 2. El P valor es la mitad del que sería si H_A fuera no direccional.

Precaución Nótese que la Tabla 7, para el test de signos y la Tabla 4, para el test t , se organizan de forma diferente. A la Tabla 7 se entra con n_D , mientras que a la Tabla 4 se entra con $(g-1)$.

Tratamiento de los ceros Puede ser que algunas de las diferencias $(Y_1 - Y_2)$ sean iguales a cero. ¿Deberían ser contadas como positivas como negativas para determinar B_s ? Un procedimiento recomendado es descartar del análisis las parejas correspondientes y reducir el tamaño de la muestra n_D de forma acorde. En otras palabras, cada pareja cuya diferencia sea cero se ignora completamente. Se considera que esas parejas no proporcionan evidencia en contra de H_0 en ninguna dirección. Nótese que este procedimiento no tiene ninguno paralelo en el test t . El test t trata las diferencias de cero de la misma forma que cualquier otro valor.

Ejemplo 8.4.3 Distribución nula

Considere un experimento con 10 parejas, de forma que $n_D = 10$. Si H_0 es cierta, entonces la distribución de probabilidad de N_+ es una distribución binomial con $n = 10$ y $p = 0,5$. La Figura 8.4.1(a) muestra esta distribución binomial, junto con los valores asociados de N_+ y N_- . La Figura 8.4.1(b) muestra la distribución nula de B_s , que es una versión «plegada» de la Figura 8.4.1(a). (Ya vimos una relación similar entre las partes (a) y (b) de la Figura 7.10.4).

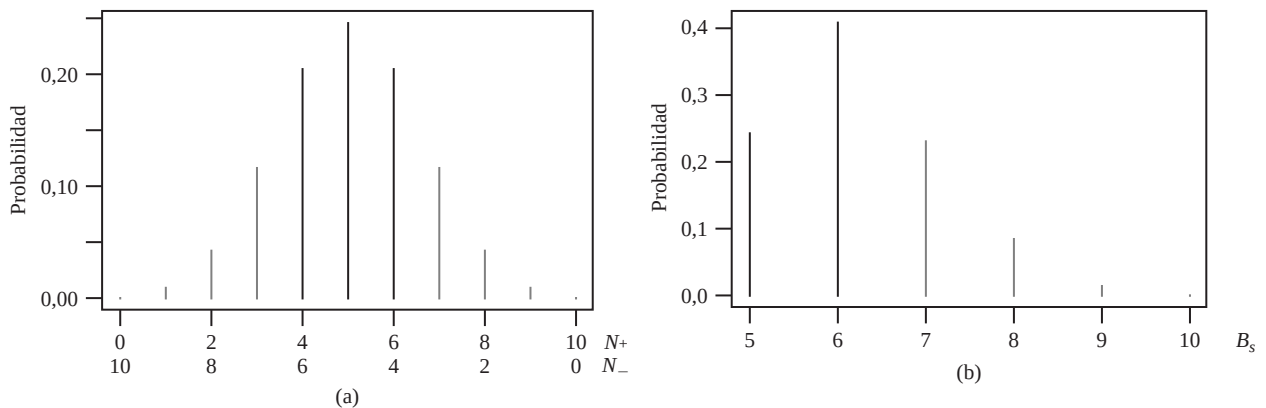


Figura 8.4.1 Distribuciones nulas para el test de signos cuando $n_D = 10$. (a) Distribución de N_+ y N_- y (b) distribución de B_s

Si N_+ es 7 y H_A es direccional (y predice que las diferencias positivas son más probables que las diferencias negativas), entonces el P valor es la probabilidad de siete o más signos (+) en 10 pruebas. Utilizando la fórmula binomial del Capítulo 3, o un computador, obtenemos que esta probabilidad es 0,17188*. Este valor (0,17188) es la suma de las barras sombreadas en la cola derecha de la Figura 8.4.1(a). Si H_A es no direccional, entonces el P valor es la suma de las barras sombreadas de la cola izquierda y de la cola derecha de la Figura 8.4.1(a). Las dos áreas sombreadas son ambas iguales a 0,17188. En consecuencia, el área total sombreada, que es el P valor, es

$$P = 2(0,17188) = 0,34376 \approx 0,34$$

En términos de la distribución nula de B_s , el P valor es la probabilidad de una cola superior. Por tanto, la suma de las barras sombreadas en la Figura 8.4.1(b) es igual a 0,34. ■

Cómo se calcula la Tabla 7 A lo largo de nuestro estudio sobre estadística nos podemos preguntar cómo tener fe en los valores críticos dados en varias tablas. La Tabla 7 es una excepción. El ejemplo siguiente muestra cómo podríamos (si quisiéramos) calcular los valores críticos nosotros mismos. Entender el ejemplo nos ayudará a apreciar cómo se han obtenido las otras tablas de valores críticos.

Ejemplo 8.4.4

Supongamos que $n_D = 10$. Vimos en el Ejemplo 8.4.3 que

Si $B_s = 7$, el P valor de los datos es 0,34376.

Cálculos similares utilizando la fórmula binomial demuestran que

Si $B_s = 8$, el P valor de los datos es 0,10938.

Si $B_s = 9$, el P valor de los datos es 0,02148.

Si $B_s = 10$, el P valor de los datos es 0,00195.

Para $n_D = 10$, los valores críticos de la Tabla 7 se reproducen en la Tabla 8.4.4.

Tabla 8.4.4 Valores críticos y P valores para el test de signos cuando $n_D = 10$							
n_D	0,20	0,10	0,05	0,02	0,01	0,002	0,001
19	8 0,109	9 0,021	9 0,021	10 0,002	10 0,002	10 0,0020	

El mínimo valor de B_s que da un P valor menor que 0,20 es $B_s = 8$, por lo que esta es la entrada en la columna de 0,20. Para $\alpha = 0,10$ o $\alpha = 0,05$, se necesita $B_s = 9$. La posibilidad más extrema, $B_s = 10$, da un P valor de 0,00195, que se redondea a 0,0020 en la tabla. No es posible obtener un P valor no direccional tan pequeño como 0,001, por lo que esa entrada se deja en blanco. ■

Aplicabilidad del test de signos

El test de signos es válido en cualquier situación en donde los valores D sean independientes entre sí y la hipótesis nula se pueda traducir apropiadamente como

$$H_0: \Pr\{D \text{ sea positiva}\} = 0,5$$

Por tanto, el test de signos es independiente de la distribución. Su validez no depende de ninguna condición sobre la forma de la distribución poblacional de los D . Esta amplia validez tiene un precio: si la distribución poblacional de los D es de hecho normal, entonces el test de signos es mucho menos potente que el test t .

El test de signos es útil porque se puede aplicar rápidamente en una amplia variedad de situaciones. De hecho, algunas veces el test de signos se puede aplicar a datos que no permiten un test t , como se mostró en el Ejemplo 8.4.1. Existe

* Aplicando la fórmula binomial tenemos

$${}_{10}C_7(0,5)^7(0,5)^3 + {}_{10}C_8(0,5)^8(0,5)^2 + {}_{10}C_9(0,5)^9(0,5)^1 + {}_{10}C_{10}(0,5)^{10} = 0,11719 + 0,04394 + 0,00977 + 0,00098 = 0,17188$$

otro test para datos emparejados, el test de signos con rango de Wilcoxon, que se presentará en la Sección 8.5, que es generalmente más potente que el test de signos y sigue siendo independiente de la distribución. Sin embargo, el test de rangos con signo de Wilcoxon es más difícil que realizar el test de signos y, como en el caso del test t , existen situaciones en las que no se puede realizar. A continuación se presenta otro ejemplo en el cual solo es posible un test de signos.

Ejemplo 8.4.5 THC y quimioterapia

La quimioterapia para el cáncer produce a menudo náuseas y vómitos. Se comparó la efectividad del THC (tetrahidrocannabinol, el ingrediente activo de la marihuana) en la prevención de estos efectos colaterales, con la efectividad de la medicina estándar Compazina. De los 46 pacientes que probaron ambos medicamentos (pero no se les dijo cuál era cuál), 21 no expresaron preferencia, mientras que 20 prefirieron el THC y 5 prefirieron la Compazina. Como «preferencia» indica un signo de la diferencia, pero no una magnitud, en esta situación es imposible realizar un test t . En el caso de un test de signos, tenemos $n_D = 25$ y $B_s = 20$, de forma que el P valor es 0,004. Incluso con $\alpha = 0,005$ podríamos rechazar $\alpha = 0,05$ y encontrar que los datos proporcionan evidencia suficiente para concluir que el THC es preferido a la Compazina¹⁴. ■

Ejercicios 8.4.1-8.4.11

8.4.1 Utilice la Tabla 7 para obtener el P valor de un test de signos (contra una alternativa no direccional) asumiendo que $n_D = 25$ y

- (a) $B_s = 6$ (b) $B_s = 7$ (c) $B_s = 8$ (d) $B_s = 9$

8.4.2 Utilice la Tabla 7 para obtener el P valor de un test de signos (contra una alternativa no direccional) asumiendo que $n_D = 15$ y

- (a) $B_s = 10$ (b) $B_s = 11$ (c) $B_s = 12$
(d) $B_s = 13$ (e) $B_s = 14$ (f) $B_s = 15$

8.4.3 Un grupo de 30 mujeres menopáusicas recibieron estrógenos conjugados orales durante un mes. Los niveles en plasma del inhibidor tipo 1 del activador-plasminógeno (IAP-1) bajaron en 22 de las mujeres, pero subieron en 8 mujeres¹⁵. Utilice un test de signos para contrastar la hipótesis nula de que los estrógenos conjugados orales no tienen efecto sobre los niveles de IAP-1. Utilice $\alpha = 0,10$ y una alternativa no direccional.

8.4.4 ¿Puede el ejercicio mental crear «músculo mental»? En un estudio sobre esta cuestión se utilizaron 12 parejas de ratas jóvenes de la misma camada. Un miembro de la pareja, elegido aleatoriamente, se crió en un entorno «enriquecido» con juguetes y compañeras mientras que su compañera se crió sola en un entorno «empobrecido». Después de 80 días se sacrificaron los animales y sus cerebros fueron diseccionados por un investigador que no conocía qué tratamiento había recibido cada rata. Una variable de interés era el peso del córtex cerebral, expresado en relación al peso total del cerebro. En 10 de las 12 parejas, el peso relativo del córtex fue mayor en la rata «enriquecida» que en su compañera «empobrecida». En las otras dos parejas, la rata «empobrecida» tenía un córtex más grande. Utilice un test de signos para comparar los entornos con $\alpha = 0,05$. Como la hipótesis alternativa considere que el entorno enriquecido tiende a incrementar el tamaño relativo del córtex¹⁶.

8.4.5 Veinte pacientes epilépticos institucionalizados participaron en un estudio de un nuevo medicamento anticonvulsivo, valproato. Diez de los pacientes (escogidos aleatoriamente), comenzaron a tomar valproato diariamente y los 10 restantes recibieron una píldora placebo idéntica. Durante un periodo de observación de ocho semanas, se apuntó el número de ataques epilépticos mayores y menores de cada paciente. Después, todos los pacientes se «cruzaron» al otro tratamiento y se apuntaron los ataques durante un segundo periodo de observación de ocho semanas. Los números de ataques menores se muestran en la tabla siguiente¹⁷. Contraste la eficiencia del valproato utilizando el test de signos con $\alpha = 0,05$. Utilice una alternativa direccional. (Nótese que este análisis ignora el posible efecto del tiempo, es decir, el primer periodo de observación frente al segundo).

Número de paciente	Periodo del placebo	Periodo del Valproato	Número de paciente	Periodo del placebo	Periodo del Valproato
1	37	5	11	7	8
2	52	22	12	9	8
3	63	41	13	65	30
4	2	4	14	52	22
5	25	32	15	6	11
6	29	20	16	17	1
7	15	10	17	54	31
8	52	25	18	27	15
9	19	17	19	36	13
10	12	14	20	5	5

8.4.6 Un investigador en ecología estudió la interacción entre los pájaros de dos subespecies, el Junco de Carolina y

el Junco del Norte. Puso juntos en un aviario un macho de Carolina y un macho del Norte, igualados por tamaño, y observó su comportamiento durante 45 minutos empezando al amanecer. Esto se repitió en diferentes días con diferentes parejas de pájaros. La tabla siguiente muestra el número de episodios en los que un pájaro fue dominante sobre el otro (por ejemplo, persiguiéndolo o desplazándolo de su percha)¹⁸. Utilice un test de signos para comparar las subespecies. Utilice una alternativa no direccional y $\alpha = 0,01$.

Pareja	Número de episodios en los que	
	el del Norte fue dominante	Carolina fue dominante
1	0	9
2	0	6
3	0	22
4	2	16
5	0	17
6	2	33
7	1	24
8	0	40

8.4.7

- (a) Suponga un conjunto de datos emparejados con $n_D = 4$ y $B_s = 4$. Calcule el P valor exacto de los datos cuando son analizados mediante el test de signos (contra una alternativa no direccional).
- (b) Explique por qué, en la Tabla 7 con $n_D = 3$, no aparecen valores críticos en ninguna columna.

8.4.8 Suponga que un conjunto de datos emparejados tiene $n_D = 4$. Calcule el P valor exacto de los datos cuando son analizados mediante el test de signos (contra una alternativa no direccional) si $B_s = 15$.

8.4.9 El estudio descrito en el Ejemplo 8.2.4, sobre el compuesto mCFP, incluía un grupo de hombres. Se preguntó a los hombres que valoraran el hambre que sentían al final de cada periodo de dos semanas y se calcularon las diferencias (valoración del hambre cuando se tomó mCFP, valoración del hambre cuando se tomó el placebo). La distribución de las diferencias no era normal. No obstante, se puede realizar un test de signos utilizando la siguiente información: de los ocho hombres que proporcionaron valoraciones del hambre, tres indicaron un hambre mayor con mCFP que con el placebo y cinco indicaron menos hambre con mCFP que con el placebo². Realice un test de signos con nivel $\alpha = 0,05$. Utilice una alternativa no direccional.

8.4.10 En referencia al Ejercicio 8.4.9, calcule el P valor exacto de los datos cuando son analizados mediante el test de signos. (Nota: H_A es no direccional).

8.4.11 (Potencia) Un investigador está planeando realizar un experimento para comparar dos tratamientos en los que parejas igualadas de sujetos recibirán dos tratamientos y se utilizará un test de signos, con una alternativa no direccional, para analizar la diferencia en las respuestas.

Suponga que el investigador cree que un tratamiento siempre resultará mejor que el otro. ¿Cuántas parejas necesitará tener en el experimento si desea ser capaz de rechazar H_0 cuando $\alpha = 0,05$? Si un tratamiento «gana» en todas las parejas, ¿cuál será el P valor del test resultante?

8.5 El test de rangos con signo de Wilcoxon

El **test de rangos con signo de Wilcoxon**, como el test de signos, es un método no paramétrico que se puede utilizar para comparar muestras pareadas. Realizar un test de rangos con signo de Wilcoxon es algo más complicado que realizar un test de signos, pero el test de Wilcoxon es más potente que el test de signos. Como el test de signos, el test de rangos con signo de Wilcoxon *no* requiere que los datos sean una muestra de una población distribuida normalmente.

El test de rangos con signo de Wilcoxon se basa en el conjunto de diferencias, $D = Y_1 - Y_2$. Combina la idea principal del test de signos («observar los signos de las diferencias») con la idea principal del test t de muestras pareadas («observar las magnitudes de las diferencias»).

Método

El test de rangos con signo de Wilcoxon procede en varios pasos, que presentaremos aquí en el contexto de un ejemplo.

Ejemplo 8.5.1 Densidad de células nerviosas

Un veterinario anatomista midió la densidad de células nerviosas en lugares específicos del intestino de nueve caballos. Los resultados del lugar I (región media del yeyuno) y del lugar II (región mesentérica del yeyuno) se muestran en la tabla siguiente¹⁹. Cada valor de densidad es el promedio de los valores del número de células nerviosas en cinco secciones iguales de tejido. La hipótesis nula de interés es que en la población de todos los caballos no hay diferencia entre los dos lugares.

1. El primer paso en el test de rangos con signo de Wilcoxon es calcular las diferencias, como se muestra en la Tabla 8.5.1.

Tabla 8.5.1 Densidad de células nerviosas en cada uno de los dos lugares			
Animal	Lugar I	Lugar II	Diferencia
1	50,6	38,0	12,6
2	39,2	18,6	20,6
3	35,2	23,2	12,0
4	17,0	19,0	-2,0
5	11,2	6,6	4,6
6	14,2	16,4	-2,2
7	24,2	14,4	9,8
8	37,4	37,6	-0,2
9	35,2	24,4	10,8

2. Seguidamente se obtiene el valor absoluto de cada diferencia.
3. Asignamos después un número de rango a estos valores absolutos, de menor a mayor, como se muestra en la Tabla 8.5.2.

Tabla 8.5.2			
Animal	Diferencia, d	$ d $	Rango de $ d $
1	12,6	12,6	8
2	20,6	20,6	9
3	12,0	12,0	7
4	-2,0	2,0	2
5	4,6	4,6	4
6	-2,2	2,2	3
7	9,8	9,8	5
8	-0,2	0,2	1
9	10,8	10,8	6

4. Seguidamente recuperamos los signos + y - de las diferencias en valor absoluto para producir rangos con signo, como se muestra en la Tabla 8.5.3.

Tabla 8.5.3			
Animal	Diferencia, d	Rango de $ d $	Rango con signo
1	12,6	8	8
2	20,6	9	9
3	12,0	7	7
4	-2,0	2	-2
5	4,6	4	4
6	-2,2	3	-3
7	9,8	5	5
8	-0,2	1	-1
9	10,8	6	6

5. Se suman los rangos de signo positivo para obtener W_+ . Se suman los valores absolutos de los rangos de signo negativo para obtener W_- . Para los datos de células nerviosas, $W_+ = 8 + 9 + 7 + 4 + 5 + 6 = 39$ y $W_- = 2 + 3 + 1 = 6$. El estadístico de contraste, W_s , se define como

$$W_s = \text{Máximo de } W_+ \text{ y } W_-$$

Para el caso de los datos de células nerviosas, $W_s = 39$.

6. Para obtener el P valor, se consulta la Tabla 8 (al final del libro). Parte de la Tabla 8 se reproduce en la Tabla 8.5.4.

Tabla 8.5.4 Valores críticos del test de rangos con signo de Wilcoxon cuando $n_D = 9$							
n	0,20	0,10	0,05	0,02	0,01	0,002	0,001
9	35 0,164	37 0,098	40 0,039	42 0,020	44 0,0078		

Utilizando la Tabla 8.5.4, podemos ver que para $W_s = 37$ el P valor es 0,098. Hay una evidencia débil pero sugerente ($P = 0,098$) de que existe una diferencia entre la densidad de células nerviosas en las dos regiones. (Rechazamos si es 0,10 o mayor). ■

Acotación del P valor De la misma forma que el test de signos, el test de rangos con signo de Wilcoxon tiene una distribución nula discreta. Ciertas entradas de valores críticos de la Tabla 8 están en blanco. Esta situación es familiar a partir del nuestro estudio del test de Wilcoxon-Mann-Whitney y del test de signos. Por ejemplo, si $n_D = 9$, entonces la evidencia más fuerte posible contra H_0 ocurre cuando todas las 9 diferencias son positivas (o cuando todas las 9 diferencias son negativas), en cuyo caso $W_s = 45$. Pero la probabilidad de que W_s sea igual a 45 cuando es cierta es $(1/2)^9 + (1/2)^9$, que es aproximadamente 0,0039. Por tanto, no es posible tener P valor de dos colas menor que 0,002, ni que hablar de 0,001. Eso es por lo que las dos últimas entradas están en blanco en la fila $n_D = 9$ de la Tabla 8. Nótese también que si $W_s = 45$, por ejemplo, entonces la tabla solamente nos diría que $P > 0,20$.

Alternativa direccional Para utilizar la Tabla 8 si la hipótesis alternativa es direccional, procederemos con el procedimiento habitual en dos pasos:

Paso 1. Comprobar la direccionalidad (ver si los datos se desvían con respecto a H_0 en la dirección especificada por H_A).

- (a) Si no es así, el P valor es mayor que 0,50.
- (b) Si es así, continuar con el paso 2.

Paso 2. El P valor es la mitad del que sería si H_A fuera no direccional.

Tratamiento de los ceros Si alguna de las diferencias ($Y_1 - Y_2$) es cero, entonces esos puntos de datos se borran y el tamaño de la muestra se reduce de forma acorde. Por ejemplo, si una de las nueve diferencias del Ejemplo 8.5.1 hubiera sido cero, habríamos borrado ese punto al realizar el test de Wilcoxon, de modo que el tamaño de la muestra hubiera sido de 8.

Tratamiento de los empates Si hay empates entre los valores absolutos de las diferencias (en el paso 3) se promedian los rangos de los valores empatados. Si hay empates, entonces el P valor dado por el test de rangos con signo de Wilcoxon es solamente aproximado.

Aplicabilidad del test de rangos con signo de Wilcoxon

El test de rangos con signo de Wilcoxon se puede utilizar en cualquier situación en donde los valores de D sean independientes entre sí y provengan de una distribución simétrica. No es necesario que la distribución sea normal*. La hipótesis nula es «no efecto del tratamiento» o «no diferencia de las poblaciones» y se puede expresar como

$$H_0: \mu_D = 0$$

* Estrictamente hablando, la distribución debe ser continua, lo que significa que la probabilidad de un empate es cero.

Algunas veces el test de rangos con signo de Wilcoxon se puede utilizar incluso con información incompleta. Por ejemplo, es posible realizar un test de Wilcoxon para los datos de injerto de piel del Ejemplo 8.4.1. Es cierto que no se puede calcular un valor exacto de d para dos de los pacientes, pero para ambos pacientes la diferencia es positiva y es mayor que cualquiera de las diferencias negativas. Los datos de la Tabla 8.5.5 muestran que solo hay dos diferencias negativas. La más pequeña de ellas es -1 , para el paciente 11. Esta es la mínima diferencia en valor absoluto, por lo que su rango con signo es -1 . La única con rango con signo negativo es para el paciente 7. Todos los otros rangos con signo son positivos. (El resto de este ejemplo se deja como ejercicio).

Tabla 8.5.5 Tiempos de supervivencia de trasplantes de piel			
Paciente	Compatibilidad A-LH		Signo de $d = y_1 - y_2$
	Alta y_1	Baja y_2	
1	37	29	8
2	19	13	6
3	57 +	15	42 +
4	93	26	67
5	16	11	5
6	23	18	5
7	20	26	-6
8	63	43	20
9	29	18	11
10	60 +	42	18 +
11	18	19	-1

Como con el test de Wilcoxon-Mann-Whitney para muestras independientes, hay un procedimiento asociado con el test de rangos con signo de Wilcoxon que se puede utilizar para construir un intervalo de confianza para μ_D . El procedimiento está más allá del alcance de este libro.

En resumen, cuando se trata con datos emparejados tenemos tres procedimientos de inferencia: el test t de muestras pareadas, el test de rangos con signo de Wilcoxon y el test de signos. El test t requiere que los datos procedan de una población distribuida normalmente. Si esta condición se cumple, entonces el test t es el recomendado, y es más potente que el test de Wilcoxon o el test de signos. El test de Wilcoxon no requiere normalidad, pero requiere que las diferencias provengan de una distribución simétrica y que se puedan ordenar. Tiene más potencia que el test de signos. El test de signos es el menos potente de los dos métodos, pero el más ampliamente aplicable, ya que solo requiere determinar si cada diferencia es positiva o negativa.

Ejercicios 8.5.1-8.5.7

8.5.1 Utilice la Tabla 8 para obtener el P valor de un test de rangos con signo de Wilcoxon (contra una alternativa no direccional), suponiendo que $n_D = 7$ y

- (a) $W_s = 22$ (b) $W_s = 25$
 (c) $W_s = 26$ (d) $W_s = 28$

8.5.2 Utilice la Tabla 8 para obtener el P valor de un test de rangos con signo de Wilcoxon (contra una alternativa no direccional), suponiendo que $n_D = 12$ y

- (a) $W_s = 55$ (b) $W_s = 65$
 (c) $W_s = 71$ (d) $W_s = 73$

8.5.3 El estudio descrito en el Ejemplo 8.2.4, sobre el compuesto mCFP, incluía un grupo de hombres. Se preguntó a los hombres que valoraran el hambre que sentían al final de cada periodo de dos semanas y se calcularon las diferencias (valoración del hambre cuando se tomó mCFP - valoración del hambre cuando se tomó el placebo). No están disponibles los datos para uno de los sujetos. Los datos para los otros ocho sujetos se muestran en la tabla siguiente². Analice estos datos con un test de rangos con signo de Wilcoxon con un nivel $\alpha = 0,10$. Utilice una alternativa no direccional.

Sujeto	Valoración del hambre		
	mCFP y_1	Placebo y_2	Diferencia $d = y_1 - y_2$
1	64	69	-5
2	119	112	7
3	0	28	-28
4	48	95	-47
5	65	145	-80
6	119	112	7
7	149	141	8
8	NA	NA	NA
9	99	119	-20

8.5.4 Como parte del estudio descrito en el Ejemplo 8.2.4 (y en el Ejercicio 8.5.3), sobre el compuesto mCFP, se midió el cambio de peso de nueve hombres. Se tomaron dos medidas para cada hombre: el cambio de peso cuando tomaba mCFP y el cambio de peso cuando tomaba el placebo. Los datos se muestran en la tabla siguiente². Analice estos datos con un test de rangos con signo de Wilcoxon con nivel $\alpha = 0,05$. Utilice una alternativa no direccional.

Sujeto	Cambio de peso		
	mCFP y_1	Placebo y_2	Diferencia $d = y_1 - y_2$
1	0,0	-1,1	1,1
2	-1,1	0,5	-1,6
3	-1,6	0,5	-2,1
4	-0,3	0,0	-0,3
5	-1,1	-0,5	-0,6
6	-0,9	1,3	-2,2
7	-0,5	-1,4	0,9
8	0,7	0,0	0,7
9	-1,2	-0,8	-0,4

8.5.5 Considere los datos de injerto de piel del Ejemplo 8.4.1. La Tabla 8.5.5, al final de la Sección 8.5, muestra los primeros pasos en la realización de un test de rango con signos de Wilcoxon son con hipótesis nula que la compatibilidad del A-LH no tiene efecto en el tiempo de supervivencia del injerto. Complete este test. Use $\alpha = 0,05$ y la alternativa direccional de que el tiempo de supervivencia tiende a ser mayor cuando el valor de compatibilidad es alto.

8.5.6 En una investigación sobre los posibles daños en el cerebro debidos al alcoholismo, se utilizó un procedimiento de exploración con rayos X conocido como tomografía por computador (TC) para medir las densidades cerebrales de 11 alcohólicos crónicos. Para cada uno de los

alcohólicos, se seleccionó un control que se igualó con el alcohólico en edad, sexo, educación y otros factores. Las medidas de densidad cerebral de los alcohólicos y de los controles igualados se presentan en la tabla siguiente²⁰. Utilice un test de rangos con signo de Wilcoxon para contrastar la hipótesis nula de que no existe diferencia frente la alternativa de que el alcoholismo reduce la densidad cerebral. Utilice $\alpha = 0,01$.

Pareja	Alcohólico	Control	Diferencia
1	40,1	41,3	-1,2
2	38,5	40,2	-1,7
3	36,9	37,4	-0,5
4	41,4	46,1	-4,7
5	40,6	43,9	-3,3
6	42,3	41,9	0,4
7	37,2	39,9	-2,7
8	38,6	40,4	-1,8
9	38,5	38,6	-0,1
10	38,4	38,1	0,3
11	38,1	39,5	-1,4
Media	39,14	40,66	-1,52
DT	1,72	2,56	1,58

8.5.7 El estudio descrito en el Ejemplo 8.1.1, sobre el efecto de la cafeína en el flujo sanguíneo del miocardio, tenía otra componente en la que se midió el flujo sanguíneo de 10 sujetos antes y después de consumir cafeína, pero bajo condiciones ambientales diferentes de las de los sujetos del Ejemplo 8.1.1²¹. En este caso las diferencias no siguen una distribución normal y, por tanto, un test t no sería válido. Utilice un test de rangos con signo de Wilcoxon para contrastar la hipótesis nula de que no existe diferencia frente a la alternativa de que la cafeína tiene un efecto en el flujo sanguíneo del miocardio. Utilice $\alpha = 0,01$.

Sujeto	Baselina	Cafeína	Diferencia
1	3,43	2,72	0,71
2	3,08	2,94	0,14
3	3,07	1,76	1,31
4	2,65	2,16	0,49
5	2,49	2	0,49
6	2,33	2,37	-0,04
7	2,31	2,35	-0,04
8	2,24	2,26	-0,02
9	2,17	1,72	0,45
10	1,34	1,22	0,12
Media	2,51	2,15	0,36
DT	0,59	0,50	0,43

8.6 Perspectiva

En esta sección consideraremos algunas limitaciones del análisis de datos emparejados.

Estudios antes-después

Muchos experimentos en las Ciencias de la Vida comparan medidas antes y después de alguna intervención experimental, y pueden presentar otra limitación. Estos estudios pueden ser difíciles de interpretar, porque el efecto de la intervención experimental se puede confundir con otros cambios con el tiempo. Por ejemplo, en el Ejemplo 8.2.3 obtuvimos evidencia significativa a favor de una disminución en el flujo sanguíneo en el miocardio después de tomar cafeína, pero indicamos que es posible que el flujo sanguíneo hubiera disminuido con el paso del tiempo incluso si los sujetos no hubieran tomado cafeína. Una forma de protegerse contra esta dificultad es utilizar controles concurrentes aleatorizados, como el siguiente ejemplo.

Ejemplo 8.6.1 Biorrealimentación y presión sanguínea

Un equipo médico investigó la efectividad de un programa de entrenamiento de biorrealimentación diseñado para reducir la presión sanguínea alta. Unos voluntarios se asignaron aleatoriamente al grupo de biorrealimentación o a un grupo de control. Todos los voluntarios recibieron literatura educativa sobre la salud y una breve clase. Además, el grupo de biorrealimentación recibió ocho semanas de entrenamiento de relajación, ayudada por biorrealimentación, meditación y ejercicios de respiración. Los resultados para la presión sanguínea sistólica, antes y después de las ocho semanas, se muestran en la Tabla 8.6.1²².

Tabla 8.6.1 Resultados del experimento de biorrealimentación					
Grupo	n	Presión sanguínea sistólica (mmHg)			
		Antes	Después	Diferencia	
				Media	ET
Biorrealimentación	99	145,2	131,4	13,8	1,34
Control	93	144,2	140,2	4,0	1,30

Analicemos los cambios antes-después mediante un test t para muestras pareadas con $\alpha = 0,05$. En el grupo de biorrealimentación, la presión sanguínea sistólica media fue de 13,8 mmHg. Para evaluar la significación estadística de esta disminución, el estadístico de contraste es

$$t_s = \frac{13,8}{1,34} = 10,3$$

que es altamente significativo (P valor $< 0,0001$). Sin embargo, este resultado por sí solo no demuestra la efectividad del entrenamiento mediante biorrealimentación. La disminución en la presión sanguínea podría ser, en todo o en parte, debida a otros factores, como la literatura educativa sobre la salud o la especial atención recibida por todos los participantes. De hecho, el test t para muestras pareadas aplicado al grupo de control dio como resultado

$$t_s = \frac{4,0}{1,30} = 3,08 \quad 0,001 < P \text{ valor} < 0,01$$

Por tanto, las personas que *no* recibieron entrenamiento de biorrealimentación *también* experimentaron un descenso significativo en su presión sanguínea.

Para aislar el efecto del entrenamiento de biorrealimentación, podemos comparar la experiencia de los dos grupos de tratamiento, utilizando un test t de muestras independientes *sobre las dos muestras de diferencias*. Elegimos de nuevo $\alpha = 0,05$. La diferencia entre los cambios de las medias en los dos grupos es

$$13,8 - 4,0 = 9,8 \text{ mmHg}$$

y el error típico de esta diferencia es

$$\sqrt{1,34^2 + 1,30^2} = 1,87$$

Por tanto, el estadístico t es

$$t_s = \frac{9,8}{1,87} = 5,24$$

Este contraste proporciona una evidencia fuerte ($P < 0,0001$) a favor de que el programa de biorrealimentación es efectivo. Si el diseño experimental no hubiera incluido el grupo de control, entonces esta última comparación crucial no habría sido posible, y el fundamento de la eficacia de la biorrealimentación habría sido de hecho poco firme. ■

Al analizar datos reales, es importante tener en mente los métodos estadísticos que estamos considerando responden solo a cuestiones limitadas.

El test t de muestras pareadas está limitado de dos formas:

1. Está limitado a cuestiones respecto a $\bar{D}$.
2. Está limitado a cuestiones sobre diferencias *conjuntas*.

La segunda limitación es muy amplia. Se aplica no solo a los métodos de este capítulo, sino también a los del Capítulo 7 y a muchas otras técnicas estadísticas elementales. Comentaremos separadamente estas dos limitaciones.

Limitación de $\bar{D}$

Una limitación simple del test t y del intervalo de confianza de muestras pareadas, pero que muchas veces se pasa por alto: cuando algunos de los valores de D son positivos y otros negativos, la magnitud de $\bar{D}$ no refleja la magnitud «típica» de los valores de D . El siguiente ejemplo ilustra lo engañoso que puede ser $\bar{D}$.

Ejemplo 8.6.2 Medida de colesterol en suero

Suponga que un químico clínico desea comparar dos métodos de medir el colesterol en suero. Está interesado en cómo se parecen los dos métodos entre sí. Para ello toma un espécimen de sangre de 400 pacientes, separa cada espécimen en dos partes iguales, y aplica en una parte el método A y en la otra el método B. La Tabla 8.6.2 muestra datos ficticios, exagerados para clarificar el tema.

Tabla 8.6.2 Colesterol en suero			
Especimen	Método A	Método B	$d = A - B$
1	200	234	-34
2	284	272	+12
3	146	153	-7
4	263	250	+13
5	258	232	+26
⋮	⋮	⋮	⋮
400	176	190	-14
Media	215,2	214,5	0,7
DT	45,6	59,8	18,8

En la Tabla 8.6.2, la diferencia de medias muestrales es pequeña ($\bar{d} = 0,7$). Además, los datos indican que la diferencia de las medias poblacionales es pequeña (un intervalo de confianza del 95 % es $-1,1 \text{ mg/dl} < \mu_D < 2,5 \text{ mg/dl}$). Pero estos comentarios sobre $\bar{D}$ o μ_D no consideran la cuestión central, que es: ¿cómo se parecen los métodos? De hecho, la Tabla 8.6.2 indica que los dos métodos no se parecen mucho. Las diferencias individuales entre el

método A y el método B no son pequeñas *en magnitud*. La media es pequeña porque las diferencias positivas y negativas tienden a cancelarse entre sí. Una gráfica similar a la de la Figura 8.3.1 sería muy útil para determinar visualmente cómo se parecen los métodos. Podríamos examinar una gráfica de ese tipo para ver lo cerca que los puntos se agrupan alrededor de la recta $y = x$, así como para ver la dispersión en el diagrama de caja de las diferencias. Para realizar una evaluación numérica del parecido entre los métodos no deberíamos enfocarnos en la diferencia media, $\bar{D}$. Sería mucho más relevante analizar las magnitudes absolutas (sin signo) de los valores de d (es decir, 34, 12, 7, 13, 26, etc.). Estas magnitudes se podrían analizar de varias formas: podríamos promediarlas, podríamos contar cuántas son «grandes» (por ejemplo, más de 10 mg/dl), etc. ■

Limitación del punto de vista conjunto

Consideremos un experimento emparejado en el que dos tratamientos, por ejemplo A y B, se aplican a la misma persona. Si aplicamos un test t , un test de signos, o un test de rangos con signo de Wilcoxon, estamos viendo a las personas de forma conjunta en vez de individualmente. Esto es adecuado si deseamos asumir que la diferencia (si existe) entre A y B es en una dirección consistente para todas las personas (o, al menos, que las características importantes de la diferencia se conservan incluso cuando las personas se ven *en conjunto*). El siguiente ejemplo ilustra este aspecto.

Ejemplo 8.6.3 Tratamiento del acné

Consideremos un estudio clínico para comparar dos lociones médicas para tratar el acné. Participan 20 pacientes. Cada paciente utiliza la loción A en un lado de su cara (elegido aleatoriamente) y la loción B en el otro lado. Después de tres semanas, se puntúa la mejora total de cada lado de la cara.

Supongamos primero que el lado A mejora más que el lado B en 10 pacientes, mientras que en los otros 10 el lado B mejora más. Utilizando un test de signos, este resultado está en perfecto acuerdo con la hipótesis nula. Y aun así, dos interpretaciones muy diferentes son lógicamente posibles:

Interpretación 1: los tratamientos A y B son de hecho completamente equivalentes. Su acción es indistinguible. Las diferencias observadas entre los lados A y B de la cara se deben enteramente a la variación por azar.

Interpretación 2: los tratamientos A y B son de hecho completamente diferentes. Para algunas personas (aproximadamente el 50 % de la población), el tratamiento A es más efectivo que el tratamiento B, mientras que en la mitad restante de la población el tratamiento B es más efectivo. Las diferencias observadas entre los lados A y B de la cara tienen sentido biológico*.

La misma ambigüedad de interpretación surge si los resultados favorecen a un tratamiento con respecto al otro. Por ejemplo, supongamos que el lado A mejora más que el lado B en 18 de los 20 casos, mientras que el lado B resulta favorecido en 2 pacientes. Este resultado, que es estadísticamente significativo ($P < 0,001$), podría de nuevo ser interpretado de dos formas. Podría significar que el tratamiento A es de hecho superior al B para todo el mundo, pero que la variación por azar oscureció esta superioridad en dos pacientes; o podría significar que A es superior al B en mayor parte de la gente, pero para aproximadamente el 10 % de la población ($2/10 = 0,10$) el tratamiento B es superior al A. ■

La dificultad que ilustra el Ejemplo 8.6.3 no está limitada a experimentos con parejas aleatorias. De hecho, es particularmente clara en otro tipo de experimento emparejado: la medida de cambio con el tiempo. Consideremos, por ejemplo, los datos de presión sanguínea del Ejemplo 8.6.1. Nuestra discusión sobre ese estudio se basó en una medida conjunta de la presión sanguínea: la media. Si las presiones de algunos pacientes subieran como resultado de la biorrealimentación y las de otros bajaran, estos detalles se compensarían en el análisis basado en la t de Student. Sólo se analiza el cambio promedio.

Las dificultades descritas previamente no están tampoco limitadas únicamente a experimentos con personas. Supongamos, por ejemplo, que dos fertilizantes, A y B, se van a comparar en un experimento agrónomo de campo utilizando un

* Esto puede parecer rebuscado, pero ocurren fenómenos de esta clase. Como un ejemplo obvio, considere la respuesta de pacientes a transfusiones de sangre de tipo A o de tipo B.

diseño emparejado, y que los datos se van a analizar utilizando un test t de muestras pareadas. Si el tratamiento A es superior al B en suelos ácidos, pero B es mejor que A en suelos alcalinos, este hecho podría oscurecer el resultado en un experimento que incluya suelos de ambos tipos.

El aspecto resaltado en los ejemplos anteriores es muy general. Los métodos estadísticos simples como el test de signos y el test t se diseñan para evaluar efectos de tratamientos *de forma global* (es decir, *conjuntamente*) para una población de personas, de ratones o de parcelas de tierra. La acumulación de efectos diferenciales de los tratamientos en subpoblaciones requiere una gestión más delicada, tanto en el diseño como en el análisis.

Este confinamiento al punto de vista conjunto se aplica al Capítulo 7 (muestras independientes) incluso más fuertemente que al presente capítulo. Por ejemplo, si se da el tratamiento A a un grupo de ratones y el tratamiento B a otro, es casi imposible saber cómo habría respondido un ratón del grupo A si hubiera recibido el tratamiento B. La única comparación posible es la conjunta. En la Sección 7.11 indicamos que la comparación estadística de muestras independientes depende de un «supuesto implícito»: esencialmente, el supuesto es que el fenómeno bajo estudio se pueda percibir adecuadamente desde un punto de vista global.

En muchas, quizá demasiadas, investigaciones biológicas los fenómenos de interés son razonablemente universales, por lo que este aspecto de sumergir al individuo en el conjunto no causa serios problemas. No obstante, no se debería perder de vista el hecho de que la globalización puede oscurecer importantes detalles individuales.

Informes de datos

Al comunicar los resultados experimentales, es deseable escoger un tipo de informe que transmita la información extra proporcionada por el emparejamiento. Con muestras pequeñas, se pueden utilizar gráficos, como en la Figura 8.1.1, donde los segmentos de recta proporcionan una evidencia visual clara de que el flujo sanguíneo decrece en cada sujeto.

Al publicar informes de investigación biológica, la información crucial relacionada con el emparejamiento a menudo se omite. Por ejemplo, una práctica común es presentar las medias y las desviaciones típicas de Y_1 e Y_2 pero ¡omitir la desviación típica de la diferencia, D ! Esto es un serio error. Es mejor presentar alguna descripción de D , utilizando una gráfica como la Figura 8.1.1, un histograma de los valores de D o al menos la desviación típica de D .

Ejercicios 8.6.1-8.6.4

8.6.1 33 hombres con valores altos de colesterol en suero, todos bebedores regulares de café, participaron en un estudio para ver si abstenerse de tomar café afectaría a sus niveles de colesterol. 25 de los hombres (elegidos aleatoriamente) no bebieron café durante cinco semanas, mientras que los restantes ocho hombres bebieron café en su forma habitual. La tabla que se presenta a continuación muestra los niveles de colesterol en suero (en mg/dl) en el punto inicial (al principio del estudio) y el cambio desde el punto inicial tras cinco semanas²³.

	Sin café ($n = 25$)		Café habitual ($n = 8$)	
	Media	DT	Media	DT
Punto inicial	341	37	331	30
Cambio desde el punto inicial	-35	27	+26	56

Para los siguientes tests t utilice alternativas no direccionales y $\alpha = 0,05$.

- El grupo que no bebió café experimentó un descenso de 35 mg/dl en el nivel medio de colesterol. Utilice un test t para evaluar la significación estadística de este descenso.
- El grupo que bebió café de forma habitual experimentó un aumento de 26 mg/dl en el nivel medio de colesterol. Utilice un test t para evaluar la significación estadística de este aumento.
- Utilice un test t de para comparar el cambio medio del grupo que bebió café (-35) con el cambio medio del grupo que bebió café de forma habitual (+26).

8.6.2 Ocho mujeres jóvenes participaron en un estudio para investigar la relación entre el ciclo menstrual y el consumo de alimentos. Todos los días se obtuvo mediante una entrevista la información de su dieta. El estudio fue doble ciego en el sentido de que las participantes no sabían su propósito y el entrevistador no conocía el intervalo de sus ciclos menstruales. La tabla muestra, para cada participante, el consumo calórico medio en los 10 días precedentes y los 10 días que siguieron al comienzo de su periodo menstrual (estos datos son solo para un ciclo). Prepare, para estos datos, una gráfica como la de la Figura 8.1.1²⁴.

Participante	Consumo de alimentos	
	Premenstrual	Posmenstrual
1	2.378	1.706
2	1.393	958
3	1.519	1.194
4	2.414	1.682
5	2.008	1.652
6	2.092	1.260
7	1.710	1.239
8	1.967	1.758

8.6.3 Un veterinario midió en 28 perros sanos la concentración de glucosa en la cámara anterior de su ojo izquierdo y de su ojo derecho, con los resultados que se muestran en la tabla²⁵.

Número de animal	Glucosa (mg/dl)		Número de animal	Glucosa (mg/dl)	
	Ojo derecho	Ojo izquierdo		Ojo derecho	Ojo izquierdo
1	79	79	16	80	80
2	81	82	17	78	78
3	87	91	18	112	110
4	85	86	19	89	91
5	87	92	20	87	91
6	73	74	21	71	69
7	72	74	22	92	93
8	70	66	23	91	87
9	67	67	24	102	101
10	69	69	25	116	113
11	77	78	26	84	80
12	77	77	27	78	80
13	84	83	28	94	95
14	83	82	29	100	102
15	74	75			

Utilizando un método t pareado, un intervalo de confianza del 95 % para la diferencia de medias es $-1,1 \text{ mg/dl} < \mu_D < 0,7 \text{ mg/dl}$. ¿Sugiere este resultado que, para un perro típico de la población, la diferencia de la concentración de glucosa entre los dos ojos es menor que 1,1 mg/dl? Explique su respuesta.

8.6.4 La tobramicina es un poderoso antibiótico. Para minimizar sus efectos colaterales tóxicos, la dosis se puede individualizar para cada paciente. 30 pacientes participaron en un estudio sobre la exactitud de su dosificación individualizada. Para cada paciente, se calculó la concentración de pico predicha de tobramicina en suero sanguíneo, basada en la edad, sexo, peso y otras características del paciente. Seguidamente se administró tobramicina y se midió la concentración de pico real ($\mu\text{g/ml}$). El informe contenía una tabla de resultados como la siguiente²⁶.

	Predicho	Real
Media	4,52	4,40
DT	0,90	0,85
n	30	30

¿Proporciona la tabla del informe información suficiente para juzgar si la dosificación individualizada es, en su conjunto, precisa en la predicción de la concentración de pico? Si es así, describa cómo llega a esta conclusión. Si no, describa qué información adicional se necesitaría y por qué.

Ejercicios suplementarios 8.S.1-8.S.23

8.S.1 Una voluntaria que trabajaba en un refugio de animales realizó un estudio sobre el efecto de la nébeda en los gatos del refugio. Apuntó el número de «interacciones negativas» que hizo cada uno de los 15 gatos en periodos de 15 minutos antes y después de darles una cucharadita de nébeda. Se recogieron medidas emparejadas en el mismo día dentro de esos 30 minutos. Los datos se muestran en la tabla que acompaña al ejercicio²⁷.

- Construya un intervalo de confianza del 95 % para la diferencia en el número medio de interacciones negativas.
- Construya un intervalo de confianza del 95 % de forma equivocada, utilizando el método de muestras independientes. ¿En qué se diferencia este intervalo del obtenido en el apartado (a)?

Gato	Antes (Y_1)	Después (Y_2)	Diferencia
Amelia	0	0	0
Bathsheba	3	6	-3
Boris	3	4	-1
Frank	0	1	-1
Jupiter	0	0	0
Lupine	4	5	-1
Madonna	1	3	-2
Michelangelo	2	1	1
Oregano	3	5	-2
Phantom	5	7	-2
Posh	1	0	1
Sawyer	0	1	-1
Scary	3	5	-2
Slater	0	2	-2
Tucker	2	2	0
Media	1,8	2,8	-1
DT	1,66	2,37	1,20

8.S.2 En referencia al Ejercicio 8.S.1, compare las poblaciones de antes y después utilizando un test t con $\alpha = 0,05$. Utilice una alternativa no direccional.

8.S.3 En referencia al Ejercicio 8.S.1, compare las poblaciones de antes y después utilizando un test de signos con $\alpha = 0,05$. Utilice una alternativa no direccional.

8.S.4 En referencia al Ejercicio 8.S.1, construya un diagrama de dispersión de los datos. ¿Indica el aspecto del diagrama de dispersión que el emparejamiento fue efectivo? Explique su respuesta.

8.S.5 Como parte de un estudio sobre la fisiología de la maduración del trigo, una agrónoma seleccionó seis plantas de trigo aleatoriamente de una parcela en el campo. Para cada planta, midió el contenido de humedad en dos grupos de semillas: un grupo de la parte «central» de la espiga y otro grupo de la parte «superior», con los resultados que se muestran en la tabla siguiente²⁸. Construya un intervalo de confianza del 90 % para la diferencia media del contenido de humedad de las dos regiones de la espiga.

Planta	Porcentaje de humedad	
	Parte central	Parte superior
1	62,7	59,7
2	63,6	61,6
3	60,9	58,2
4	63,0	60,5
5	62,7	60,6
6	63,7	60,8

8.S.6 Unos biólogos advirtieron que algunos peces de río se encontraban más a menudo en los remansos, que son partes profundas del río con movimiento lento del agua, mientras que otros preferían los rápidos, que son regiones poco profundas con movimiento rápido del agua. Para investigar si estos dos hábitats mantienen iguales niveles de diversidad (es decir, igual número de especies), capturaron peces en 15 localizaciones a lo largo de un río. En cada localización apuntaron el número de especies capturadas en un rápido y el número de especies capturadas en un remanso adyacente. La tabla siguiente contiene los datos²⁹. Construya un intervalo de confianza del 90 % para la diferencia en la diversidad media entre los dos tipos de hábitats.

Localización	Remanso	Rápido	Diferencia
1	6	3	3
2	6	3	3
3	3	3	0
4	8	4	4
5	5	2	3
6	2	2	0
7	6	2	4
8	7	2	5
9	1	2	-1
10	3	2	1
11	4	3	1
12	5	1	4
13	4	3	-1
14	6	2	4
15	4	3	1
Media	4,7	2,5	2,2
DT	1,91	0,74	1,86

8.S.7 En referencia al Ejercicio 8.S.6, ¿qué condiciones son necesarias para que el intervalo de confianza sea válido? ¿Se cumplen esas condiciones? ¿Cómo lo sabe?

8.S.8 En referencia al Ejercicio 8.S.6, compare los hábitats utilizando un test t con $\alpha = 0,10$. Utilice una alternativa no direccional.

8.S.9 En referencia al Ejercicio 8.S.6.

- Compare los hábitats utilizando el test de signos con $\alpha = 0,10$. Utilice una alternativa no direccional.
- Utilice la fórmula binomial para calcular el P valor exacto del apartado (a).

8.S.10 En referencia al Ejercicio 8.S.6, analice estos datos utilizando un test de rangos con signo de Wilcoxon.

8.S.11 En referencia al test de rangos con signo de Wilcoxon el Ejercicio 8.S.10, ¿con qué base se podría argumentar

que el P valor encontrado en este test podría no ser exacto? Es decir, ¿por qué se podría argumentar que el P valor del test de Wilcoxon no es una medida completamente exacta de la fuerza de la evidencia contra H_0 en este caso?

8.S.12 En un estudio sobre el efecto de la cafeína en el metabolismo muscular, nueve voluntarios varones se sometieron a pruebas de ejercicio de brazo en dos ocasiones separadas. En una ocasión, el voluntario tomó una cápsula de placebo una hora antes del test t . En la otra ocasión recibió una cápsula que contenía cafeína pura. (El orden de aplicación de las dos cápsulas se determinó aleatoriamente). Durante cada prueba de ejercicio se midió la razón de intercambio respiratorio (RER) del sujeto. La RER es la razón entre el dióxido de carbono producido y el oxígeno consumido, e indica si la energía se está obteniendo a partir de hidratos de carbono o a partir de grasa. Los resultados se presentan en la tabla siguiente³⁰. Utilice un test t para evaluar el efecto de la cafeína. Use una alternativa no direccional y $\alpha = 0,05$.

Sujeto	RER (%)	
	Placebo	Cafeína
1	105	96
2	119	99
3	92	89
4	97	95
5	96	88
6	101	95
7	94	88
8	95	93
9	98	88

8.S.13 Para los datos del Ejercicio 8.S.12, construya una gráfica como la de la Figura 8.1.1.

8.S.14 En referencia al Ejercicio 8.S.12, analice los datos utilizando un test de signos.

8.S.15 Ciertos tipos de células nerviosas tienen la propiedad de regenerar una parte de las mismas que han sido amputadas. En un estudio inicial de este proceso, se tomaron medidas de los nervios en la espina dorsal de monos *rhesus*. Los nervios que salían de la parte izquierda de la espina dorsal se cortaron, mientras que los nervios que salían de la parte derecha se dejaron intactos. Durante el proceso de regeneración, se midió el contenido de creatina fosfato (CF) en la parte izquierda y en la parte derecha de la espina dorsal. La tabla siguiente muestra los datos para el lado derecho (control, Y_1), y para el lado izquierdo (regeneración, Y_2). Las unidades de medida son mg de CF por cada 100 g de tejido³¹. Utilice un test t de para comparar los dos lados con $\alpha = 0,05$. Utilice una alternativa no direccional.

Animal	Lado derecho (control)	Lado izquierdo (regeneración)	Diferencia
1	16,3	11,5	4,8
2	4,8	3,6	1,2
3	10,9	12,5	-1,6
4	14,2	6,3	7,9
5	16,3	15,2	1,1
6	9,9	8,1	1,8
7	29,2	16,6	12,6
8	22,4	13,1	9,3
Media	15,50	10,86	4,64
DT	7,61	4,49	4,89

8.S.16 La aldosterona es una hormona que interviene en el mantenimiento del equilibrio de fluidos en el cuerpo. En un estudio veterinario se trató a seis perros con fallo cardiaco con el medicamento Captopril, y se midió la concentración en plasma de aldosterona antes y después del tratamiento. Los resultados se muestran en la tabla siguiente³². Utilice un test de signos con $\alpha = 0,10$ y una alternativa no direccional, para investigar la afirmación de que el Captopril afecta al nivel de aldosterona.

Animal	Antes	Después	Diferencia
1	749	374	375
2	469	300	169
3	343	146	197
4	314	134	180
5	286	69	217
6	223	20	203
Media	397,3	173,8	223,5
DT	190,5	136,4	76,1

8.S.17 En referencia al Ejercicio 8.S.16, analice estos datos utilizando un test de rangos con signo de Wilcoxon.

8.S.18 En referencia al Ejercicio 8.S.16, nótese que los perros de este estudio no se comparan con un grupo de control. ¿Cómo debilita esto cualquier interferencia que se pudiera hacer sobre la efectividad del Captopril?

8.S.19 (Ejercicio para computador) En una investigación sobre el mecanismo de curación de las heridas, una bióloga eligió un diseño emparejado, utilizando los cuartos traseros izquierdo y derecho de la salamandra *Notophthalmus viridescens*. Tras amputar cada cuarto trasero, hizo una pequeña herida en la piel y después mantuvo el cuarto durante cuatro horas en una solución que contenía benzamil o en una solución de control. Teorizó que el benzamil perjudicaría a la cicatrización. La tabla siguiente muestra la cantidad de cicatri-

zación, expresada como el área (mm^2) cubierta con nueva piel después de cuatro horas³³.

Animal	Cuarto de control	Cuarto de benzamil	Animal	Cuarto de control	Cuarto de benzamil
1	0,55	0,14	10	0,42	0,21
2	0,15	0,08	11	0,49	0,11
3	0,00	0,00	12	0,08	0,03
4	0,13	0,13	13	0,32	0,14
5	0,26	0,10	14	0,18	0,37
6	0,07	0,08	15	0,35	0,25
7	0,20	0,11	16	0,03	0,05
8	0,16	0,00	17	0,24	0,16
9	0,03	0,05			

- Evalúe el efecto del benzamil utilizando un test t con $\alpha = 0,05$. Considere la hipótesis alternativa que la expectativa de la investigadora es correcta.
- Proceda como en el apartado (a), pero utilice un test de signos.
- Construya un intervalo de confianza del 95 % para el efecto medio del benzamil.
- Construya un diagrama de dispersión de los datos. ¿Indica la apariencia del diagrama de dispersión que el emparejamiento fue efectivo? Explique su respuesta.

8.S.20 (Ejercicio para computador) En un estudio sobre la sugestión hipnótica, 16 voluntarios varones se asignaron aleatoriamente a un grupo experimental y a un grupo de control. Cada sujeto participó en una sesión experimental de dos fases. En la primera fase se midió su respiración mientras el sujeto estaba despierto y en reposo. (Estas medidas se describen también en los Ejercicios 7.5.6 y 7.10.4). En la segunda fase, se dijo al sujeto que imaginara que estaba realizando un trabajo muscular, y se midió de nuevo su respiración.

Los sujetos del grupo experimental fueron inducidos mediante hipnosis entre la primera y la segunda fase. Por tanto, la sugestión de imaginar el trabajo muscular fue una «sugestión hipnótica» en los sujetos experimentales y una «sugestión en estado despierto» en los sujetos de control. La tabla que acompaña a este ejercicio muestra las medidas de la ventilación total (litros de aire por minuto por metro cuadrado de área corporal) para los 16 sujetos³⁴.

- Utilice un test t para comparar los valores medios en reposo de los dos grupos. Utilice una alternativa no direccional y $\alpha = 0,05$. Esto es lo mismo que en el Ejercicio 7.5.6 (a).
- Utilice tests t pareados y no pareados adecuados para investigar (i) la respuesta del grupo experimental a la sugestión; (ii) la respuesta del grupo de control a la suges-

Grupo experimental			Grupo de control		
Sujeto	Reposo	Trabajo	Sujeto	Reposo	Trabajo
1	5,74	6,24	9	6,21	5,50
2	6,79	9,07	10	4,50	4,64
3	5,32	7,77	11	4,86	4,61
4	7,18	16,46	12	4,78	3,78
5	5,60	6,95	13	4,79	5,41
6	6,06	8,14	14	5,70	5,32
7	6,32	11,72	15	5,41	4,54
8	6,34	8,06	16	6,08	5,98

tión; (iii) la diferencia entre las respuestas del grupo experimental y del grupo de control. Utilice alternativas direccionales (la sugestión incrementa la ventilación, y la sugestión hipnótica la incrementa más que la sugestión en estado despierto) y emplee $\alpha = 0,05$ para cada test t .

- Repita la investigación del apartado (b) utilizando contrastes no paramétricos adecuados (tests de signos y de Wilcoxon-Mann-Whitney).
- Utilice gráficas adecuadas para investigar si es razonable asumir la condición de normalidad subyacente en los tests t del apartado (b). ¿Cómo arroja esta investigación luz en las discrepancias entre los resultados de los apartados (b) y (c)?

8.S.21 Suponga que se desea contrastar si un medicamento experimental reduce la presión sanguínea más de lo que lo hace un placebo. Se planifica administrar el medicamento o el placebo a algunos sujetos y apuntar cuanto se reduce su presión sanguínea. Se tienen 20 sujetos disponibles.

- Se podrían formar 10 parejas igualadas, formando cada pareja igualando sujetos, lo mejor que se pueda, basándose en la edad y el sexo y después asignar aleatoriamente un sujeto de cada pareja al medicamento y el otro sujeto de la pareja al placebo. Explique por qué utilizar un diseño de parejas igualadas sería una buena idea.
- Explique brevemente por qué un diseño de parejas igualadas no sería una buena idea. Es decir, ¿de qué forma podría un diseño de ese tipo ser inferior a un diseño completamente aleatorizado?

8.S.22 Un grupo de 20 mujeres menopáusicas recibieron estradiol intradérmico durante un mes. Los niveles en plasma del inhibidor del activador del plasminógeno tipo 1 (IAP-1) disminuyeron en 10 de las mujeres y aumentaron en las otras 10³⁵. Utilice un test de signos para contrastar la hipótesis nula de que el estradiol intradérmico no tiene efecto en el nivel del IAP-1. Utilice $\alpha = 0,05$ y una alternativa no direccional.

8.5.23 Seis pacientes con enfermedad renal recibieron plasmáfesis. Se midió la excreción de proteína urinaria (gramos de proteína por gramo de creatinina) antes y después de la plasmáfesis. Los datos se muestran en la tabla siguiente³⁶. Utilice estos datos para investigar si la plasmáfesis afecta o no a la excreción de proteína urinaria en los pacientes con enfermedad renal. (*Sugerencia*: haga una gráfica de los datos y considere si es apropiado un test t en la escala original).

Paciente	Antes	Después	Diferencia
1	20,3	0,8	19,5
2	9,3	0,1	9,2
3	7,6	3,0	4,6
4	6,1	0,6	5,5
5	5,8	0,9	4,9
6	4,0	0,2	3,8
Media	8,9	0,9	7,9
DT	5,9	1,1	6,0

Notas

1. Namdar, M., Koepfli, P., Grathwohl, R., Siegrist, P. T., Klainguti, M., Schepis, T., Delaloye, R., Wyss, C. A., Fleischmann, S. P., Gaemperli, O. y Kaufmann, P. A. (2006). Caffeine decreases exercise-induced myocardial flow reserve. *Journal of the American College of Cardiology* **47**, 405-410.

2. Sargent, P. A., Sharpley, A. L., Williams, C., Goodall, E. M. y Cowen, P. J. (1997). 5-HT_{2C} receptor activation decreases appetite and body weight in obese subjects. *Psychopharmacology* **133**, 309-312. Las tasas de apetito se midieron «en escalas analíticas visuales de 10 cm».

3. Datos no publicados cortesía de R. Buchman. Los datos se tomaron en Oberlin, Ohio, durante la primavera de 2001.

4. Day, K. M., Patterson, F. L., Luetkemeier, O. W., Ohm, H. W., Polizotto, K., Roberts, J. J., Shaner, G. E., Huber, D. M., Finney, R. E., Foster, J. E. y Gallun, R. L. (1980). Performance and adaptation of small grains in Indiana. *Station Bulletin*, n.º 290. West Lafayette, Ind.: Agricultural Experiment Station of Purdue University. Datos iniciales cortesía de W. E. Nyquist. La prueba real incluyó más de dos variedades.

5. Datos no publicados cortesía de C. H. Noller.

6. Cicirelli, M. F. y Smith, L. D. (1985). Cyclic AMP levels during the maturation of *Xeno pus* oocytes. *Developmental Biology* **108**, 254-258. Datos iniciales cortesía de M. F. Cicirelli.

7. Judge, M. D., Aberle, E. D., Cross, H. R. y Schanbacher, B. D. (1984). Thermal shrinkage temperature of intramuscular collagen of bulls and steers. *Journal of Animal Science* **59**, 706-709. Datos iniciales cortesía de los autores y E. W. Mills.

8. Swedo, S. E., Leonard, H. L., Rapoport, J. L., Lenane, M. C., Goldberger, E. L. y Cheslow, B. S. (1989). A double-blind comparison of clomipramine and desipramine in the treatment of trichotillomania (hair pulling). *New England Journal of Medicine* **321**, 497-501.

9. Datos no publicados cortesía de A. Ladavac. Datos tomados en Oberlin, Ohio, en noviembre de 1996.

10. En un estudio sin emparejamiento natural (por ejemplo, si no hay disponibles gemelos idénticos), se pueden tomar 2 grupos de igual tamaño y crear parejas utilizando variables como edad y peso. Si se realiza un experimento en el que miembros de una pareja se asignan aleatoriamente a grupos opuestos de tratamiento, el análisis de datos emparejados tiene buenas propiedades. Sin embargo, si el estudio es observacional (de forma que no hay asignación aleatoria dentro de las parejas), un análisis de datos emparejados, como un test t , tenderá a subestimar la verdadera variabilidad de la diferencia que se está estudiando y la verdadera probabilidad de error de Tipo I de un test t será mayor que la nominal. Para una explicación, véase David, H. A. y Gunnink, J. L. (1997). The paired t test under artificial pairing. *The American Statistician* **51**, 9-12.

11. Schriewer, H., Guennewig, V., Assmann, G. (1983). Effect of 10 weeks endurance training on the concentration of lipids and lipoproteins as well as on the composition of high-density lipoproteins in blood serum. *International Journal of Sports Medicine* **4**, 109-115. Reimpreso con permiso de Georg Thieme Verlag KG.

12. Datos de los experimentos presentados en varios artículos, por ejemplo, Fout, G. S., Simon, E. H. (1983). Antiviral activities directed against wild-type and interferon-sensitive mengovirus. *Journal of General Virology* **64**, 1543-1555. Datos iniciales cortesía de E. H. Simon. La unidad de medida es proporcional al número de placas formadas por los virus sobre una sola capa de células de ratón. Como se obtuvieron con una técnica de disolución en serie, las medidas tienen un número variable de dígitos significativos; los ceros al final de los número de tres cifras no son dígitos significativos.

13. Adaptado de Batchelor, J. R. y Hackett, M. (1970). HL-A matching in treatment of burned patients with skin allografts. *Lancet* **2**, 581-583.

14. Sallan, S. E., Cronin, C., Zelen, M. y Zinberg, N. E. (1980). Antiemetics in patients receiving chemotherapy for cancer. *New England Journal of Medicine* **302**, 135-138. Reimpreso con permiso.
15. Koh, K. K., Mincemoyer, R., Bui, M. N., Csako, G., Pucino, F., Guetta, V., Waclawiw, M. y Cannon, R. O. (1997). Effects of hormone replacement therapy on fibrinolysis in postmenopausal women. *New England Journal of Medicine* **336**, 683-690. Datos iniciales cortesía de K. K. Koh.
16. Rosenzweig, M. R., Bennett, E. L. y Diamond, M. C. (1972). Brain changes in response to experience. *Scientific American* **226**, n.º 2, 22-29. Also Bennett, E. L., Diamond, M. C., Krech, D. y Rosenzweig, M. R. (1964). Chemical and anatomical plasticity of brain. *Science* **146**, 610-619. Copyright 1964 de la American Association for the Advancement of Science.
17. Richens, A. y Ahmad, S. (1975). Controlled trial of valproate in severe epilepsy. *British Medical Journal* **4**, 255-256.
18. Wiedenmann, R. N., and Rabenold, K. N. (1987). The effects of social dominance between two subspecies of dark-eyed juncos, *Junco hyemalis*. *Animal Behavior* **35**, 856-864. Datos iniciales cortesía de los autores.
19. Mast, J. (1983). Innervation of the equine small intestine. Master's thesis, Purdue University. Datos iniciales cortesía del autor.
20. Golden, C. J., Graber, B., Blose, I., Berg, R., Coffman, J. y Block, S. (1981). Difference in brain densities between chronic alcoholic and normal control patients. *Science* **211**, 508-510. Datos iniciales cortesía de C. J. Golden. Copyright 1981 de la AAAS.
21. Datos de Namdar, M., *et al.* El experimento descrito en el Ejemplo 8.1.1 se realizó en condiciones de gran altitud simulada, mientras que el experimento descrito en el Ejercicio 8.5.7 se realizó bajo condiciones que imitaban a las del nivel del mar.
22. Patel, C., Marmot, M. G. y Terry, D. J. (1981). Controlled trial of biofeedback-aided behavioural methods in reducing mild hypertension. *British Medical Journal* **282**, 2005-2008.
23. Forde, O. H., Knutsen, S. F., Arnesen, E. y Thelle, D. S. (1985). The Tromsø heart study: Coffee consumption and serum lipid concentrations in men with hypercholesterolaemia: A randomised intervention study. *British Medical Journal* **290**, 893-895. (Los tamaños muestrales no son iguales porque los 25 hombres que no tomaron café representaban realmente 3 grupos de tratamiento diferentes, que siguieron el mismo régimen las primeras cinco semanas del estudio y regímenes diferentes después).
24. Dalvit, S. P. (1981). The effect of the menstrual cycle on patterns of food intake. *American Journal of Clinical Nutrition* **34**, 1811-1815.
25. Datos no publicados cortesía de D. J. Honory W. A. Vestre.
26. Sesin, G. P. (1984). Pharmacokinetic dosing of Tobramycin sulfate. *American Pharmacy NS24*, 778.
27. Vakoutis, J., Stein, G. E., Miller, P. B. y Clayman, A. E. (1981). Aminoglycoside monitoring program. *American Journal of Hospital Pharmacy* **38**, 1477-1480. Copyright 1981, American Society of Hospital Pharmacists, Inc. Todos los derechos reservados. Reimpreso con permiso.
28. Jovan, S. (2000). Catnip bonanza. *Stats*, n.º 27, 25-27.
29. Dale, E. M. y Housley, T. L. (1986). Sucrose synthase activity in developing wheat endosperms differing in maximum weight. *Plant Physiology* **82**, 7-10. Datos iniciales cortesía de los autores.
30. Dale, E. M. y Housley, T. L. (1986). Sucrose synthase activity in developing wheat endosperms differing in maximum weight. *Plant Physiology* **82**, 7-10. Datos iniciales cortesía de los autores.
31. Datos no publicados cortesía de M. Heithaus y D. Rogers. Las muestras se tomaron del Vermilion River al norte de Ohio durante la primavera de 1995.
32. Salib, N. M. (1985). The effect of caffeine on the respiratory exchange ratio of separate submaximal arms and legs exercise of middle distance runners. Master's thesis, Purdue University.
33. Adaptado de Bodian, D. (1947). Nucleic acid in nerve-cell regeneration. *Symposia of the Society for Experimental Biology*, n.º 1, *Nucleic Acid*, 163-178. Utilizado con permiso de The Society for Experimental Biology.
34. Knowlen, G. G., Kittleson, M. D., Nachreiner, R. F. y Eyster, G. E. (1983). Comparison of plasma aldosterone concentration among clinical status groups of dogs with chronic heart failure. *Journal of the American Veterinary Medical Association* **183**, 991-996.
35. Robinson, L. R. (1985). The effects of electrical fields on wound healing in *Notophthalmus viridescens*. Master's thesis, Purdue University. Datos iniciales cortesía del autor y de J. W. Vanable, Jr.
36. Agosti, E. y Camerota, G. (1965). Some effects of hypnotic suggestion on respiratory function. *International Journal of Clinical and Experimental Hypnosis* **13**, 149-156. El experimento incluyó realmente una tercera fase.
37. Koh, K. K., Mincemoyer, R., Bui, M. N., Csako, G., Pucino, F., Guetta, V., Waclawiw, M. y Cannon, R. O. (1997). Effects of hormone replacement therapy on fibrinolysis in postmenopausal women. *New England Journal of Medicine* **336**, 683-690. Datos iniciales cortesía de K. K. Koh.
38. Savin, V. J., Sharma, R., Sharma, M., McCarthy, E. T., Swan, S. K., Ellis, E., Lovell, H., Warady, B., Gunwar, S., Chonko, A. M., Artero, M. y Vincenti, F. (1996). Circulating factor associated with increased glomerular permeability to albumin in recurrent focal segmental glomerulosclerosis. *New Engl and Journal of Medicine* **334**, 878-883. Datos iniciales cortesía de V. J. Savin.

DATOS EN CATEGORÍAS: DISTRIBUCIONES PARA UNA MUESTRA

9

Objetivos

En este capítulo estudiaremos los datos en categorías. Concretamente:

- exploraremos las distribuciones muestrales de los estimadores que describen poblaciones dicotómicas;
- demostraremos cómo calcular e interpretar intervalos de confianza para proporciones;
- proporcionaremos un método para obtener el tamaño óptimo de la muestra para estimar una proporción;
- demostraremos cómo y cuándo realizar un test chi-cuadrado de bondad de ajuste.

9.1 Observaciones dicotómicas

En el Capítulo 5 trabajamos con problemas que involucraban variables numéricas y examinamos la distribución muestral de la media. En el Capítulo 6 utilizamos la distribución muestral para explicar cómo la media muestral tiende a variar con respecto a la media poblacional y construimos intervalos de confianza para la media poblacional. Empezaremos este capítulo procediendo de forma similar y considerando primero una variable categórica dicotómica (es decir, una variable categórica que solo tiene dos posibles valores) y la distribución muestral de la proporción. En la Sección 9.2 utilizaremos la distribución muestral de la proporción muestral para construir un intervalo de confianza para una proporción poblacional.

Proporción muestral ajustada de Wilson, $\tilde{P}$

Cuando se muestrea una población dicotómica grande, un estimador natural de la proporción poblacional, p , es la proporción muestral, $\hat{p} = y/n$, siendo y el número de observaciones de la muestra que verifican el atributo de interés y n el tamaño de la muestra.

Ejemplo 9.1.1 Soda contaminada

En un cierto momento, los dispensadores de refrescos pueden albergar bacterias como *Chryseobacterium meningosepticum*, que pueden causar enfermedades¹. Para estimar la proporción de dispensadores de refrescos contaminados en una comunidad de Virginia, unos investigadores muestrearon aleatoriamente 30 dispensadores y encontraron que 5 estaban contaminados con *Chryseobacterium meningosepticum*. Por tanto, la proporción muestral de dispensadores contaminados es

$$\hat{p} = \frac{5}{30} = 0,167$$

■

El valor del estimador, $\hat{p} = 0,167$, dado en el Ejemplo 9.1.1 es una buena estimación de la proporción poblacional de dispensadores de refrescos contaminados, pero no es la única estimación posible. La proporción muestral ajustada de Wilson, $\tilde{P}$, es otro estimador de la proporción poblacional y está dada por la fórmula siguiente.

Proporción muestral ajustada de Wilson, $\tilde{P}$

$$\tilde{p} = \frac{y + 2}{n + 4}$$

Ejemplo 9.1.2 Soda contaminada

La proporción muestral ajustada de Wilson de dispensadores contaminados es

$$\tilde{p} = \frac{5 + 2}{30 + 4} = 0,206^*$$

Como ilustra el ejemplo anterior, $\tilde{P}$ es equivalente a calcular la proporción muestral ordinaria $\hat{P}$ de una muestra aumentada: una que incluye cuatro observaciones extra de dispensadores de refrescos, dos que están contaminados y dos que no. Este aumento tiene el efecto de predisponer la estimación hacia el valor 1/2. Hablando en términos generales, deberíamos evitar estimadores sesgados, pero como veremos en la Sección 9.2, los intervalos de confianza basados en este estimador sesgado, $\tilde{P}$, son realmente más fiables que los basados en $\hat{P}$.

La distribución de muestreo de $\tilde{P}$

En el Capítulo 3 vimos cómo utilizar la distribución binomial para calcular las probabilidades de todas las posibles composiciones de muestra con la finalidad de realizar un muestreo aleatorio a partir de una gran población dicotómica. Estas probabilidades, a su vez, determinan la distribución de muestreo de $\tilde{P}$, como muestra el siguiente ejemplo.

Ejemplo 9.1.3 Soda contaminada

Suponga que en una cierta región de Estados Unidos el 17 % de todos los dispensadores de refrescos están contaminados con *Chryseobacterium meningosepticum*. Si examináramos una muestra aleatoria de dos dispensadores de refrescos de esta población de dispensadores, entonces tendríamos cero, una o dos máquinas contaminadas. La probabilidad de que ambos dispensadores estén contaminados es $0,17 \times 0,17 = 0,0289$. La probabilidad de que ninguno esté contaminado es $(1 - 0,17) \times (1 - 0,17) = 0,6889$. Hay dos formas de obtener una muestra en la que una máquina está contaminada y la otra no. La primera podría estar contaminada, pero no la segunda, o viceversa. Por tanto, la probabilidad de que exactamente una máquina esté contaminada es

$$0,17 \times (1 - 0,17) + 0,17 \times (1 - 0,17) = 0,2822$$

Si $\tilde{P}$ representa la proporción muestral ajustada de Wilson de dispensadores contaminados, entonces una muestra que no contenga dispensadores contaminados tiene un valor de $\tilde{p} = \frac{0 + 2}{2 + 4} = 0,33$, lo que ocurre con probabilidad 0,6889. Una muestra que contenga una máquina contaminada tiene un valor de $\tilde{p} = \frac{1 + 2}{2 + 4} = 0,50$, lo que sucede con probabilidad 0,2822. Finalmente, una muestra que contenga dos máquinas contaminadas tiene un valor de $\tilde{p} = \frac{2 + 2}{2 + 4} = 0,67$, lo que sucede con probabilidad 0,0289[†]. Por tanto, hay aproximadamente un 69 % de probabilidad de que $\tilde{P}$ sea igual a 0,33, un 28 % de probabilidad de que $\tilde{P}$ sea igual a 0,50 y un 3 % de probabilidad de que $\tilde{P}$ sea igual a 0,67.

Esta distribución muestral se presenta en la Tabla 9.1.1 y en la Figura 9.1.1.

* Manteniendo nuestro convenio, $\tilde{P}$ indica una variable aleatoria, mientras que $\tilde{p}$ indica un número concreto (como 0,206 en este ejemplo).

[†] Merece la pena comentar que con un tamaño de muestra pequeño ($n = 2$) los posibles valores de $\tilde{p}$ son 0,33, 0,50 y 0,67, mientras los posibles valores de $\hat{p}$ son 0,00, 0,50 y 1,00. Esto arroja luz sobre por qué $\tilde{p}$ es un estimador sensible a la proporción poblacional, particularmente para muestras pequeñas. Con una muestra pequeña es bastante probable que no se obtuvieran máquinas contaminadas incluso aunque una proporción razonable de la población estuviera contaminada. Sería poco prudente, con una muestra tan pequeña, asegurar que la proporción poblacional de máquinas contaminadas es 0.

Tabla 9.1.1 Distribución muestral de Y (el número de dispensadores contaminados) y de $\tilde{P}$ (la proporción ajustada de Wilson de dispensadores contaminados) para muestras de tamaño $n = 2$ en una población con el 17 % de los dispensadores contaminados

Y	$\tilde{P}$	Probabilidad
0	0,33	0,6889
1	0,50	0,2822
2	0,67	0,0289

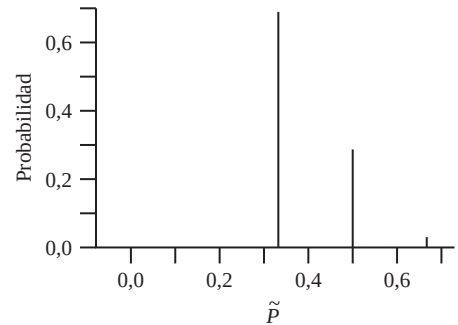


Figura 9.1.1 Distribución del muestreo de $\tilde{P}$ para $n = 2$ y $p = 0,17$

Ejemplo 9.1.4 Soda contaminada

Supongamos que deseamos examinar una muestra de 20 dispensadores de una población en la que el 17 % están contaminados. ¿Cuántos dispensadores contaminados deberíamos esperar obtener en la muestra? Como se vio en el Ejemplo 9.1.3, esta cuestión se puede responder en el lenguaje de la probabilidad. Sin embargo, como $n = 20$ es más bien grande, no haremos una lista con todas las posibles muestras. En vez de eso, haremos los cálculos utilizando la distribución binomial con $n = 20$ y $p = 0,17$. Por ejemplo, calculemos la probabilidad de que cinco dispensadores de la muestra estén contaminados y 15 no:

$$\begin{aligned}\Pr\{5 \text{ contaminados, } 15 \text{ no contaminados}\} &= {}_{20}C_5(0,17)^5(0,83)^{15} \\ &= 15.504(0,17)^5(0,83)^{15} \\ &= 0,1345\end{aligned}$$

Siendo $\tilde{P}$ la proporción muestral ajustada de Wilson de dispensadores contaminados, una muestra que contenga 5 dispensadores contaminados tiene $\tilde{p} = \frac{5+2}{20+4} = 0,2917$. Por tanto, hemos obtenido que

$$\Pr\{\tilde{P} = 0,2917\} = 0,1345$$

Se puede utilizar la distribución binomial para determinar la distribución muestral completa de $\tilde{P}$. Esta distribución se muestra en la Tabla 9.1.2 y su histograma de probabilidades en la Figura 9.1.2.

Tabla 9.1.2 Distribución al muestreo de Y , el número de éxitos, y de $\tilde{P}$, la proporción de éxitos ajustada por Wilson, cuando $n = 20$ y $p = 0,17$

Y	$\tilde{P}$	Probabilidad	Y	$\tilde{P}$	Probabilidad
0	0,0833	0,0241	11	0,5417	0,0001
1	0,1250	0,0986	12	0,5833	0,0000
2	0,1667	0,1919	13	0,6250	0,0000
3	0,2083	0,2358	14	0,6667	0,0000
4	0,2500	0,2053	15	0,7083	0,0000
5	0,2917	0,1345	16	0,7500	0,0000
6	0,3333	0,0689	17	0,7917	0,0000
7	0,3750	0,0282	18	0,8333	0,0000
8	0,4167	0,0094	19	0,8750	0,0000
9	0,4583	0,0026	20	0,9167	0,0000
10	0,5000	0,0006			

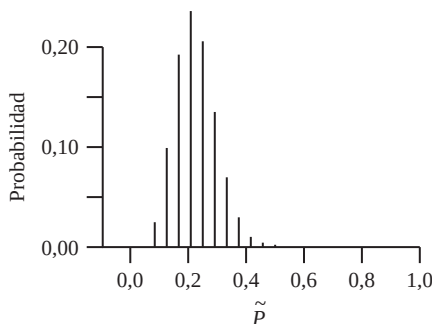


Figura 9.1.2 Distribución del muestreo de $\tilde{P}$ cuando $n = 20$ y $p = 0,17$

Podemos utilizar esta distribución para responder preguntas como «si tomamos una muestra aleatoria de tamaño $n = 20$, ¿cuál es la probabilidad de que no más de 5 estén contaminados?». Nótese que esta pregunta se puede hacer de dos formas equivalentes: «¿Cuánto vale $\Pr\{Y \leq 5\}$?» y «¿cuanto vale $\Pr\{\tilde{P} \leq 0,2917\}$?» La respuesta a las dos preguntas se obtiene sumando las primeras seis probabilidades de la Tabla 9.1.2.

$$\begin{aligned}\Pr\{Y \leq 5\} &= \Pr\{\tilde{P} \leq 0,2917\} \\ &= 0,0241 + 0,0986 + 0,1919 + 0,2358 + 0,2053 + 0,1345 \\ &= 0,8902\end{aligned}$$

■

Relación con la inferencia estadística

Al hacer una inferencia estadística a partir de una muestra de la población, es razonable utilizar $\tilde{P}$ como un estimador de p . La distribución en el muestreo de $\tilde{P}$ se puede utilizar para predecir cuánto error de muestreo se puede esperar en el estimador. Por ejemplo, supongamos que deseamos saber si el error de muestreo será menor que 5 puntos porcentuales, o en otras palabras, si $\tilde{P}$ estará dentro de un intervalo de amplitud $\pm 0,05$ alrededor de p . No podemos predecir con certeza si este suceso ocurrirá, pero podemos obtener la probabilidad de que ocurra, como se ilustra en el siguiente ejemplo.

Ejemplo 9.1.5 Soda contaminada

En el ejemplo del dispensador de soda con $n = 20$ podemos ver en la Tabla 9.1.2 que

$$\begin{aligned}\Pr\{0,12 \leq \tilde{P} \leq 0,22\} &= 0,0986 + 0,1919 + 0,2358 \\ &= 0,5263 \approx 0,53\end{aligned}$$

Por tanto, hay un 53 % de probabilidad de que, dada una muestra de tamaño 20, $\tilde{P}$ esté dentro del intervalo de amplitud 0,05 alrededor de p .

■

Dependencia con el tamaño de la muestra

De la misma forma que la distribución muestral de $\bar{Y}$ depende de n , también lo hace la distribución muestral de $\tilde{P}$. Cuanto más grande sea el valor de n , más probable será que $\tilde{P}$ esté cerca de p^* . El siguiente ejemplo ilustra este efecto.

Ejemplo 9.1.6 Soda contaminada

La Figura 9.1.3 muestra la distribución muestral de $\tilde{P}$, para tres valores diferentes de n en la población de dispensadores de refrescos del Ejemplo 9.1.1. (Cada distribución muestral está determinada por una distribución binomial con $p = 0,17$).

* Esta afirmación podría interpretarse demasiado literalmente. Como función de n , la probabilidad de que $\tilde{P}$ esté cerca de p tiene una tendencia global creciente, pero puede fluctuar algo.

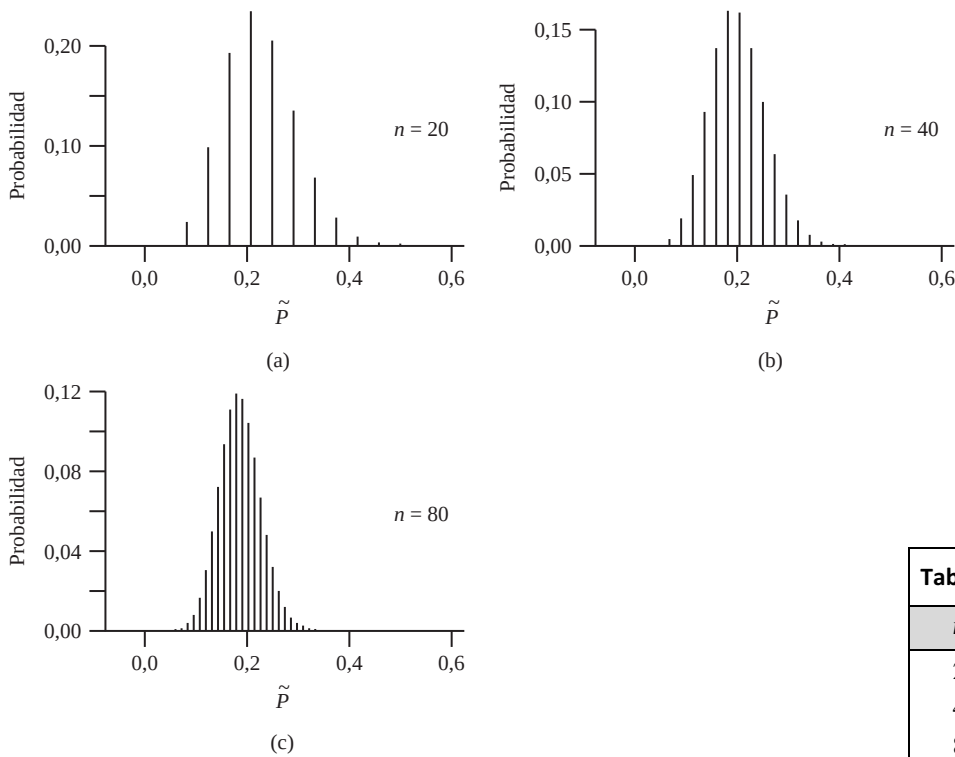


Figura 9.1.3 Distribuciones muestrales de $\tilde{P}$ para $p = 0,17$ y varios valores de n

Tabla 9.1.3

n	$\Pr\{0,12 \leq \tilde{P} \leq 0,22\}$
20	0,53
40	0,56
80	0,75
400	0,99

Puede verse en la figura que cuando n aumenta, la distribución muestral se va comprimiendo más alrededor del valor de $p = 0.17$. Por tanto, la probabilidad de que $\tilde{P}$ esté cerca de p tiende a incrementarse cuando n aumenta. Por ejemplo, consideremos la probabilidad de que $\tilde{P}$ esté dentro de ± 5 puntos porcentuales alrededor de p . Vimos en el Ejemplo 9.1.5 que para $n = 20$ esta probabilidad es igual a 0,53. La Tabla 9.1.3 y la Figura 9.1.3 muestran la dependencia de la probabilidad con n .

Nota: Una muestra mayor aumenta la probabilidad de que $\tilde{P}$ esté cerca de p . Sin embargo, deberíamos ser conscientes de que la probabilidad de que $\tilde{P}$ sea exactamente *igual* a p es muy pequeña para n grande. De hecho,

$$\Pr\{\tilde{P} = 0,17\} = 0,110 \text{ para } n = 80^*$$

El valor $\Pr\{0,12 \leq \tilde{P} \leq 0,22\} = 0,75$ es la suma de muchas probabilidades pequeñas, la mayor de las cuales vale 0,110. Este efecto se puede ver claramente en la Figura 9.1.3(c). ■

Ejercicios 9.1.1-9.1.10

9.1.1 Considere que se toma una muestra aleatoria de tamaño 3 de una población de personas que fuman y se apunta cuántas de ellas, si hay, tienen cáncer de pulmón. Sea $\tilde{P}$ la proporción ajustada de Wilson de personas de la muestra con cáncer de pulmón. ¿Cuáles son los posibles valores de la distribución muestral de $\tilde{P}$?

9.1.2 Suponga que se va a extraer una muestra aleatoria de tres individuos de una población grande en la que el 37 % de los individuos son mutantes (como en el Ejemplo 3.6.4). Considere $\tilde{P}$ la proporción ajustada de Wilson de mutantes en la muestra. Calcule la probabilidad de que $\tilde{P}$ sea igual a

(a) $2/7$ (b) $3/7$

* Para $n = 80$, $\tilde{p} = 0,1677$ cuando $y = 12$, es el valor más cercano posible a 0,17.

¿Es posible obtener una muestra de tres individuos en la que $\tilde{P}$ sea cero? Explique su respuesta.

9.1.3 Suponga que se va a extraer una muestra aleatoria de cinco individuos de una población grande en la que el 37 % de los individuos son mutantes (como en el Ejemplo 3.6.4). Sea $\tilde{P}$ la proporción ajustada de Wilson de mutantes en la muestra.

(a) Utilice los resultados de la Tabla 3.6.3 para determinar la probabilidad de que $\tilde{P}$ sea igual a

(i) 2/9	(ii) 3/9	(iii) 4/9
(iv) 5/9	(v) 6/9	(vi) 7/9

(b) Represente la distribución muestral de $\tilde{P}$ en una gráfica similar a la de la Figura 9.1.1.

9.1.4 Un nuevo tratamiento para el síndrome de inmunodeficiencia adquirida (SIDA) se va a experimentar en una pequeña prueba clínica sobre 15 pacientes. La proporción ajustada de Wilson $\tilde{P}$ de los que responden al tratamiento se utilizará como un estimador de la proporción p de (potenciales) respuestas en la población total de pacientes de SIDA. Si de hecho $p = 0,2$, y si los 15 pacientes se pueden ver como una muestra aleatoria de la población, calcule la probabilidad de que

- (a) $\tilde{P} = 5/19$ (b) $\tilde{P} = 2/19$

9.1.5 En un cierto bosque, el 25 % de los árboles de pino blanco están infectados con la roya vesicular del pino. Supongamos que se elige una muestra aleatoria de cuatro pinos blancos, y sea $\tilde{P}$ la proporción muestral ajustada de Wilson de los árboles infectados.

(a) Calcule la probabilidad de que sea $\tilde{P}$ igual a

- (i) 2/8 (ii) 3/8 (iii) 4/8 (iv) 5/8 (v) 6/8

(b) Represente la distribución muestral de $\tilde{P}$ en una gráfica similar a la Figura 9.1.1.

9.1.6 En referencia al Ejercicio 9.1.5:

(a) Determine la distribución muestral de $\tilde{P}$ para muestras de tamaño $n = 8$ árboles de pino blanco del mismo bosque.

(b) Represente gráficas de las distribuciones muestrales de $\tilde{P}$ para $n = 4$ y para $n = 8$, utilizando las mismas escalas horizontales y verticales para ambas. Compare las dos distribuciones visualmente. ¿En qué se diferencian?

9.1.7 La concha del caracol de tierra *Limocolaria marfensiana* tiene dos posibles formas de color: estriada y pálida. En una cierta población de estos caracoles, el 60 % de los individuos tienen conchas estriadas (como en el Ejercicio 3.6.4). Supongamos que se toma una muestra de 36 caracoles de la población. Sea $\tilde{P}$ la proporción muestral ajustada de Wilson de conchas estriadas. Calcule

- (a) $\Pr\{\tilde{P} = 0,5\}$ (b) $\Pr\{\tilde{P} = 0,6\}$
 (c) $\Pr\{\tilde{P} = 0,7\}$ (d) $\Pr\{0,5 \leq \tilde{P} \leq 0,7\}$
 (e) El porcentaje de muestras para las que $\tilde{P}$ está dentro de $\pm 0,10$ alrededor de p .

9.1.8 En cierta comunidad, el 17 % de las máquinas dispensadores de soda están contaminadas (como en el Ejemplo 9.1.5). Suponga que se toma una muestra aleatoria de cinco dispensadoras y se observa la contaminación. Sea $\tilde{P}$ la proporción muestral ajustada de Wilson de dispensadoras contaminadas.

- (a) Calcule la distribución muestral de $\tilde{P}$.
 (b) Represente un histograma de la distribución obtenida en el apartado (a) y compárelo visualmente con la Figura 9.1.3. ¿En qué difieren las dos distribuciones?

9.1.9 Considere el muestreo aleatorio de una población dicotómica. Sea E el suceso de que $\tilde{P}$ está dentro de $\pm 0,05$ alrededor de p . En el Ejemplo 9.1.5 encontramos que $\Pr\{E\} = 0,53$ para $n = 20$ y $p = 0,17$. Calcule $\Pr\{E\}$ para $n = 20$ y $p = 0,25$. (Quizá de forma sorprendente, las dos probabilidades son aproximadamente iguales).

9.1.10 Considere que se toma una muestra aleatoria de tamaño 10 de la población de estudiantes de una cierta universidad y se pregunta a cada uno de ellos si fuman o no. En el contexto de este ejercicio, explique lo que significa la distribución muestral de $\hat{p}$, la proporción muestral ordinaria.

9.2 Intervalo de confianza para una proporción poblacional

En la Sección 6.3 explicamos los intervalos de confianza cuando la variable observada es cuantitativa. Se pueden usar ideas similares para construir intervalos de confianza en situaciones en las que la variable es *categorica* y el parámetro de interés es una *proporción* poblacional. Suponemos que los datos se pueden considerar como una muestra aleatoria de alguna población. En esta sección explicaremos la construcción del intervalo de confianza para una proporción poblacional.

Consideremos una muestra aleatoria de n observaciones categóricas, y fijemos nuestra atención en una de las categorías. Por ejemplo, supongamos que un genetista observa n conejillos de Indias cuyo color del pelaje puede ser negro, sepia, crema o albino. Fijemos nuestra atención en la categoría «negro». Sea p la proporción poblacional de la categoría de interés, y sea $\tilde{p}$ la proporción muestral ajustada de Wilson (como en la Sección 9.1), que es nuestra estimación de p . La situación se representa esquemáticamente en la Figura 9.2.1.

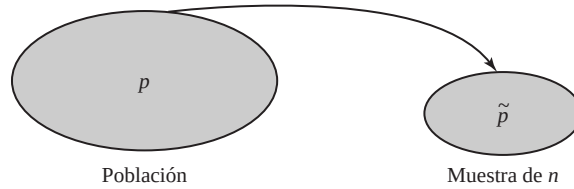


Figura 9.2.1 Notación para la población y la proporción muestral

¿Cuál es la probabilidad de que $\tilde{P}$ esté cercana a p ? Vimos en la Sección 9.1 que esta pregunta se puede responder en términos de la distribución muestral de $\tilde{P}$ (que a su vez se calcula a partir de la distribución binomial). Como veremos, utilizando propiedades de la distribución muestral de $\tilde{P}$, tales como el error típico y que los valores de $\tilde{P}$ tienen un comportamiento aproximadamente normal bajo ciertas situaciones, podremos construir intervalos de confianza para p . Para construir los intervalos utilizaremos los mismos razonamientos utilizados para datos numéricos en la Sección 6.3, donde construimos intervalos de confianza para μ basados en las propiedades de la distribución muestral de $\bar{Y}$.

Aunque se puede construir directamente un intervalo de confianza para p a partir de la distribución binomial, en muchas situaciones prácticas se puede utilizar como alternativa un método aproximado más simple. Cuando el tamaño de la muestra, n , es grande, la distribución muestral de $\tilde{P}$ es aproximadamente normal. Esta aproximación está relacionada con el Teorema Central del Límite. Si revisamos la Figura 9.1.2, veremos que las distribuciones muestrales se parecen a curvas normales, especialmente la distribución con $n = 80$. (La aproximación se explica con detalle en la Sección opcional 5.4). En la Sección 6.3 planteamos que, cuando los datos proceden de una población normal, un intervalo de confianza del 95 % para la media poblacional μ se construye como

$$\bar{y} \pm t_{0,025} \text{ET}_{\bar{Y}}$$

El intervalo de confianza para la proporción poblacional se construye de forma análoga. Utilizaremos $\tilde{P}$ como el centro de un intervalo de confianza del 95 % para p . Para proceder necesitamos calcular el error típico de $\tilde{P}$.

Error típico de $\tilde{P}$

El error típico del estimador se obtiene utilizando la siguiente fórmula.

Error típico de $\tilde{P}$ (para un intervalo de confianza del 95 %)

$$\text{ET}_{\tilde{P}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}$$

Esta fórmula del error típico del estimador es similar a la fórmula del error típico de la media, pero con $\sqrt{\tilde{p}(1 - \tilde{p})}$ jugando el papel de s y $n + 4$ en lugar de n .

Ejemplo 9.2.1 Fumar durante el embarazo

Como parte del Estudio Nacional del Crecimiento Familiar, se preguntó sobre sus hábitos de fumar a 496 mujeres con edades entre 20 y 24 años que habían dado a luz². 78 de las mujeres muestreadas indicaron que habían fumado durante el embarazo, lo que equivale al 15,7 % ($78/496 = 0,157$ o 15,7 %); por tanto, $\tilde{p}$ es $\frac{78 + 2}{496 + 4} = \frac{80}{500} = 0,16$.

El error típico es $\sqrt{\frac{0,16(1 - 0,16)}{500}} = 0,016$ o 1,6 %. Un valor muestral de $\tilde{P}$ está típicamente dentro de 2 errores típicos alrededor de la proporción poblacional p . Basándose en este error típico, podemos esperar que la proporción, p , de todas las mujeres con edades entre 20 y 24 años que fumaron durante el embarazo esté en el intervalo (0,128, 0,192) o (12,8 %, 19,2 %). El intervalo de confianza para p hace esta idea más precisa. ■

Intervalo de confianza del 95 % para p

Una vez que tenemos el error típico de $\tilde{P}$, necesitamos saber cómo es de probable que $\tilde{P}$ esté cerca de p . El proceso general de construcción del intervalo de confianza para una proporción es similar al utilizado en la Sección 6.3 para construir un intervalo de confianza para la media. Sin embargo, al construir un intervalo de confianza para la media, multiplicábamos el error típico por un multiplicador t . Esto se basaba en tomar una muestra de una distribución normal. Cuando tratamos con datos de proporciones sabemos que la población no es normal (¡solo hay dos valores en la población!), pero el Teorema Central del Límite nos dice que la distribución muestral de $\tilde{P}$ es aproximadamente normal si el tamaño de la muestra, n , es grande. Es más, sucede que incluso para muestras moderadas o pequeñas, los intervalos basados en $\tilde{P}$ y multiplicadores Z hacen muy buen trabajo al estimar la proporción poblacional, p^3 .

Para un intervalo de confianza del 95 %, el multiplicador Z apropiado es $z_{0,025} = 1,960$. Por tanto, el intervalo de confianza del 95 % aproximado para una proporción poblacional p se construye como se indica en el siguiente recuadro*.

Intervalo de confianza del 95 % para p

Intervalo de confianza del 95 %: $\tilde{p} \pm 1,96ET_{\tilde{P}}$

Ejemplo 9.2.2 Cáncer de pecho

El gen *BRCA1* se ha relacionado con el cáncer de pecho. Unos investigadores utilizaron análisis del ADN para buscar mutaciones del *BRCA1* en 169 mujeres con historiales familiares de cáncer de pecho. De las 169 mujeres, 27 (16 %) tenían mutaciones del *BRCA1*⁴. Sea p la probabilidad de que una mujer con historial familiar de cáncer de pecho tenga una mutación del *BRCA1*. Para estos datos, $\tilde{p} = \frac{27}{169} = 0,168$. El error típico de $\tilde{P}$ es

$$\sqrt{\frac{0,168(1 - 0,168)}{169 + 4}} = 0,028. \text{ Por tanto, un intervalo de confianza del 95 \% para } p \text{ es}$$

$$0,168 \pm (1,96)(0,028)$$

o

$$0,168 \pm 0,055$$

o

$$0,113 < p < 0,223$$

Por tanto, tenemos una confianza del 95 % en que la probabilidad de una mutación del *BRCA1* en una mujer con un historial familiar de cáncer de pecho esté entre 0,113 y 0,223 (es decir, entre el 11,3 % y el 22,3 %). ■

Nótese que el tamaño del error típico es inversamente proporcional a $\sqrt{n}$, como se ilustra en el siguiente ejemplo.

Ejemplo 9.2.3 Cáncer de pecho

Suponga, como en el Ejemplo 9.2.2, que una muestra de mujeres con historiales familiares de cáncer de pecho contiene el 16 % con mutaciones del *BRCA1*. Entonces, $\tilde{p} = 0,168$ y

$$ET_{\tilde{P}} \approx \sqrt{\frac{0,168(0,832)}{n + 4}}$$

* Muchos libros de estadística presentan el intervalo de confianza para la proporción como $\hat{p} \pm 1,96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$, siendo $\hat{p} = y/n$.

Este intervalo comúnmente utilizado es similar al intervalo que presentamos, particularmente si n es grande. Para tamaños de muestra pequeños o moderados, el intervalo que presentamos tiene mayor probabilidad de contener a la proporción poblacional p . En el Apéndice 9.1 se presenta una explicación técnica del intervalo de Wilson utilizando $\tilde{P}$.

Vimos en el Ejemplo 9.2.2 que si $n = 169$, entonces

$$ET_{\hat{p}} = 0,028$$

Si $n = 4 \times 169 = 676$, entonces

$$ET_{\hat{p}} = 0,014$$

Por tanto, una muestra con la misma composición (es decir, 16 % con mutaciones del *BRCA1*), pero cuatro veces mayor tendría dos veces más precisión en la estimación de p . ■

La proporción muestral ajustada por Wilson se puede utilizar para construir un intervalo de confianza para p incluso cuando el tamaño de la muestra es pequeño, como se ilustra en el siguiente ejemplo.

Ejemplo 9.2.4 OMEC

La oxigenación mediante membrana extracorpórea (OMEC) es un procedimiento para salvar la vida que se utiliza para tratar a niños recién nacidos que sufren fallo respiratorio severo. Se realizó un experimento en el que 11 bebés se trataron con OMEC. Ninguno de los 11 bebés murió⁵. Sea p la probabilidad de muerte de un bebé tratado con OMEC. El hecho de que ninguno de los bebés del experimento muriera no nos debería llevar a creer que la probabilidad de muerte, p , es exactamente cero, sino solo que es cercana a cero. El estimador dado por $\hat{p}$ es $2/15 = 0,133$. El error típico de $\hat{p}$ es

$$\sqrt{\frac{0,133(0,867)}{15}} = 0,088^*$$

Por tanto, un intervalo de confianza del 95 % para p es

$$0,133 \pm (1,96)(0,088)$$

o

$$0,133 \pm 0,172$$

o

$$-0,039 < p < 0,305$$

Sabemos que p no puede ser negativa, por lo que establecemos el intervalo de confianza como $(0, 0,305)$.

Por tanto, tenemos una confianza del 95 % en que la probabilidad de muerte de un recién nacido con fallo respiratorio severo que ha sido tratado con OMEC esté entre 0 y 0,305 (es decir, entre 0 % y 30,5 %). ■

Intervalos de confianza unilaterales

La mayoría de los intervalos de confianza son de la forma «estimación $\pm$ margen de error», y se conocen como intervalos bilaterales. Sin embargo, es posible construir un intervalo de confianza unilateral, que será apropiado cuando solo es de interés un límite inferior o un límite superior. El siguiente ejemplo proporciona una ilustración.

Ejemplo 9.2.5 OMEC-unilateral

Considere los datos de OMEC del Ejemplo 9.2.4, que se utilizan para estimar la probabilidad de muerte, p , de un recién nacido con fallo respiratorio severo. Sabemos que p no puede ser menor que cero, pero desearíamos saber cómo podría ser de grande. Mientras que un intervalo de confianza bilateral se basa en capturar la mitad de la distribución normal estándar que contiene el 95 % de probabilidad y, por tanto, utiliza los multiplicadores Z de $\pm 1,96$, un intervalo de confianza del 95 % unilateral (superior) utiliza el hecho de que $\Pr\{-\infty < Z < 1,645\} = 0,95$. Por tanto, el límite superior del intervalo de confianza es $\hat{p} + 1,645 \times ET_{\hat{p}}$ y el límite inferior del intervalo es menos infinito. En este caso tenemos

$$0,133 + (1,645)(0,088) = 0,133 + 0,145 = 0,278$$

* Nótese que si hubiéramos utilizado el método presentado comúnmente de $\hat{p} \pm 1,96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$, hubiéramos obtenido que el error típico es cero, resultando un intervalo de confianza de 0 ± 0 . ¡Un intervalo así no parece muy útil en la práctica!

como límite superior. El intervalo resultante es $(-\infty, 0,278)$, pero como p no puede ser negativa, establecemos el intervalo de confianza como $(0, 0,278)$. Es decir, tenemos una confianza del 95 % en que la probabilidad de muerte es como máximo el 27,8 %.

Planificación de un estudio para estimar p

En la Sección 6.4 explicamos un método para escoger el tamaño de la muestra n de forma que el estudio propuesto tuviera la suficiente precisión para el objetivo buscado. El planteamiento dependía de dos elementos: (1) una especificación del $ET_{\bar{y}}$ deseado y (2) una estimación preliminar de la DT. En el presente contexto, cuando la variable observada es categórica, se puede utilizar un planteamiento similar. Se especifica un valor deseado de $ET_{\tilde{p}}$, y si se dispone de una información inicial sobre $\tilde{p}$, entonces el tamaño muestral requerido n se puede determinar mediante la siguiente ecuación:

$$ET_{\text{deseado}} = \sqrt{\frac{(\text{Estimación inicial de } \tilde{p})(1 - \text{Estimación inicial de } \tilde{p})}{n + 4}}$$

El siguiente ejemplo ilustra el uso del método.

Ejemplo 9.2.6 Condición de zurdo

En un estudio sobre estudiantes universitarios ingleses y escoceses, 40 de 400 estudiantes varones eran zurdos⁶. La estimación de la proporción muestral es

$$\tilde{p} = \frac{40 + 2}{400 + 4} \approx 0,104$$

Supongamos que consideramos estos datos con un estudio piloto y que deseamos planificar un estudio lo suficientemente grande como para estimar p con un error típico de un punto porcentual, es decir, 0,01. Escogemos n de forma que se cumpla la siguiente relación:

$$\sqrt{\frac{0,104(0,896)}{n + 4}} \leq 0,01$$

Esta ecuación se resuelve fácilmente obteniéndose que $n + 4 \geq 931,8$. Deberíamos planificar una muestra de 928 estudiantes.

Planificación sin información Supongamos que no se dispone de una información preliminar de p . Aún así, en esta situación todavía es posible planificar un experimento para conseguir un valor deseado de $ET_{\tilde{p}}$ *. Este plan «ciego» depende del hecho de que la cantidad $\sqrt{\tilde{p}(1 - \tilde{p})}$ es *máxima* cuando $\tilde{p} = 0,5$. Esto se puede ver en la gráfica de la Figura 9.2.2. Esto permite deducir que un valor de n calculado utilizando «Estimación preliminar de $\tilde{p}$ » = 0,5 será *conservativa*, es decir, será lo suficientemente grande. (Por supuesto, puede ser mucho más grande que el valor necesario si $\tilde{p}$ es realmente muy diferente de 0,5). El ejemplo siguiente muestra cómo utilizar esta planificación en el «peor de los casos».

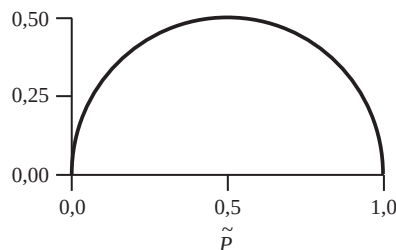


Figura 9.2.2 Dependencia con $\tilde{p}$ de $\sqrt{\tilde{p}(1 - \tilde{p})}$

* Sin embargo, esto no sería posible si estuviéramos planificando un estudio para estimar una media poblacional μ y no tuviéramos ninguna información sobre el valor del ET.

Ejemplo 9.2.7 Condición de zurdo

Supongamos, como en el Ejemplo 9.2.6, que estamos planificando un estudio sobre las personas zurdas y que deseamos que $ET_{\tilde{p}}$ sea 0,01, pero supongamos también que no tenemos ninguna información preliminar. Podemos proceder como en el Ejemplo 9.2.6, pero utilizando una estimación preliminar de $\tilde{p}$ de 0,5. Tenemos entonces

$$\sqrt{\frac{0,5(0,5)}{n+4}} \leq 0,01$$

lo que significa que $n+4 \geq 2.500$, por lo que necesitamos $n = 2.496$. Por tanto, una muestra de 2.496 estudiantes sería adecuada para estimar p con un error típico de 0,01, independientemente del valor real de p . (Por supuesto, si $p = 0,1$, este valor de n es mucho mayor que el necesario). ■

Ejercicios 9.2.1-9.2.13

9.2.1 Una serie de pacientes con infecciones bacterianas de heridas fueron tratados con el antibiótico cefotaxime. La respuesta bacteriológica (desaparición de las bacterias de la herida) fue considerada «satisfactoria» en el 84 % de los pacientes⁷. Determine el error típico de $\tilde{P}$, la proporción observada ajustada de Wilson de respuestas «satisfactorias», si la serie contuviera:

- (a) 50 pacientes de los cuales 42 fueron considerados «satisfactorios».
- (b) 200 pacientes de los cuales 168 fueron considerados «satisfactorios».

9.2.2 En un experimento con una cierta mutación de la mosca de la fruta *Drosophila* se examinaron n individuos. Se encontró que el 20 % de ellos eran mutantes. Determine el error típico de $\tilde{P}$ si

- (a) $n = 100$ (20 mutantes).
- (b) $n = 400$ (80 mutantes).

9.2.3 En referencia al Ejercicio 9.2.2, en cada caso ($n = 100$ y $n = 400$), construya un intervalo de confianza del 95 % para la proporción poblacional de mutantes.

9.2.4 En una población natural de ratones (*Mus musculus*), cerca de Ann Arbor, Michigan, el pelaje de algunos individuos tiene manchas blancas en el vientre. En una muestra de 580 ratones de la población se encontró que 28 individuos tenían manchas blancas en el vientre⁸. Construya un intervalo de confianza del 95 % para la proporción poblacional de este rasgo.

9.2.5 Para evaluar la política de vacunación rutinaria de niños contra la tos ferina, se monitorizaron las reacciones adversas en 339 niños que recibieron su primera inyección de vacuna. Se detectaron reacciones en 69 niños⁹.

- (a) Construya un intervalo de confianza del 95 % para la probabilidad de una reacción adversa a la vacuna.
- (b) Interprete el intervalo de confianza del apartado (a). ¿Qué indica el intervalo sobre las vacunaciones contra la tos ferina?

(c) Utilizando el intervalo obtenido en el apartado (a), ¿se puede tener confianza en que la probabilidad de una reacción adversa a la vacuna es menor que 0,25?

(d) ¿Qué nivel de confianza está asociado a su respuesta al apartado (c)? (Sugerencia: ¿cuál es el nivel de confianza asociado al intervalo unilateral?)

9.2.6 En un estudio sobre los tipos de sangre humana en primates no humanos se estudió una muestra de 71 orangutanes y se encontró que 14 tenían grupo sanguíneo B¹⁰. Construya un intervalo de confianza del 95 % para la frecuencia relativa de sangre de tipo B en la población de orangutanes.

9.2.7 En la población del caracol *Cepaea*, las conchas de algunos individuos tienen bandas oscuras, mientras que las de otros individuos no tienen bandas¹¹. Supongamos que una bióloga está planificando un estudio para estimar el porcentaje de individuos con bandas en una cierta población natural, y que desea estimar el porcentaje (que anticipa que será de alrededor de un 60 %) con un error típico que no exceda los 4 puntos porcentuales. ¿Cuántos caracoles debería recoger?

9.2.8 (Continuación del Ejercicio 9.2.7) ¿Cuál sería la respuesta si el porcentaje anticipado de caracoles con bandas fuera del 50 % en vez del 60 %?

9.2.9 La habilidad de reconocer con el gusto el compuesto feniltiocarbamida (FTC) es un rasgo controlado genéticamente en los humanos. En Europa y Asia, aproximadamente el 70 % de las personas son «reconocedores»¹². Suponga que se está planificando un estudio para estimar la frecuencia relativa de reconocedores en cierta población de Asia y se desea que el error típico de la frecuencia relativa estimada sea de 0,01. ¿Cuántas personas deberían incluirse en el estudio?

9.2.10 En referencia al Ejercicio 9.2.9, suponga que se está planificando un estudio para una parte del mundo en la que el porcentaje de reconocedores es totalmente desconocido, de forma que la cifra del 70 % utilizada en el Ejercicio 9.2.9

no es aplicable. ¿Qué tamaño de la muestra es necesario para que el error típico no sea mayor que 0,01?

9.2.11 En referencia al Ejercicio 9.2.9, suponga que el requisito sobre el ET se relaja en un factor de 2, es decir, de 0,01 a 0,02. ¿Reduciría esto el tamaño muestral necesario en un factor de 2? Explique su respuesta.

9.2.12 La variedad de trigo «Luso» es resistente a la mosca Hessian. Para entender el mecanismo genético que controla la resistencia, un agrónomo planea examinar la descendencia de un cierto cruce entre trigo Luso y una variedad no resistente. Cada planta de la descendencia se clasificará como resistente o susceptible y el agrónomo estimará la proporción de la descendencia que es resistente¹³. ¿Cuántas plantas de

/la descendencia necesitaría clasificar para garantizar que el error típico de la estimación de la proporción no será mayor que 0,05?

9.2.13 (Continuación del Ejercicio 9.2.12) Suponga que el agrónomo está considerando dos posibles mecanismos genéticos de herencia de la resistencia. La razón poblacional de descendientes resistentes a susceptibles sería de 1:1 bajo un mecanismo y de 3:1 bajo el otro. Si el agrónomo utiliza el tamaño de la muestra determinado en el Ejercicio 9.2.12, ¿podría asegurar que un intervalo de confianza del 95 % excluiría al menos uno de los mecanismos? Es decir, ¿podría asegurar que el intervalo de confianza *no* contendrá simultáneamente los valores 0,50 y 0,75? Explique su respuesta.

9.3 Otros niveles de confianza (opcional)

El procedimiento explicado en la Sección 9.2 se puede utilizar para construir intervalos de confianza del 95 %. Para construir intervalos con otros niveles de confianza, son necesarias algunas modificaciones al procedimiento. La primera modificación afecta a $\tilde{p}$. Para un intervalo de confianza del 95 % definimos $\tilde{p}$ como $\frac{y+2}{n+4}$. En general, para un intervalo de confianza de nivel $100(1-\alpha)\%$, $\tilde{p}$ se define como

$$\tilde{p} = \frac{y + 0,5(z_{\alpha/2}^2)}{n + z_{\alpha/2}^2}$$

Para un intervalo de confianza del 95 %, $z_{\alpha/2}$ es 1,96, por lo que $\tilde{p} = \frac{y + 0,5(1,96^2)}{n + 1,96^2}$. Esto es igual a $\frac{y + 1,92}{n + 3,84}$, que se redondea a $\frac{y+2}{n+4}$. Sin embargo, se puede usar cualquier nivel de confianza. Por ejemplo, para un intervalo de confianza del 90 %, $\tilde{p} = \frac{y + 0,5(1,645^2)}{n + 1,645^2}$, que es igual a $\frac{y + 1,35}{n + 2,7}$.

La segunda modificación afecta al error típico. Para un intervalo de confianza del 95 % utilizamos $\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$ como término del error típico. En general, se utiliza $\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+z_{\alpha/2}^2}}$ como término del error típico.

Finalmente, el multiplicador Z debe ajustarse al nivel de confianza (1,645 para un intervalo de confianza del 90 %, etc.). Esos valores se pueden obtener muy fácilmente en la Tabla 4 con $gl = \infty$. [Recuérdese de la Sección 6.3 que la distribución t con $gl = \infty$ es una distribución normal (Z)]. El ejemplo siguiente ilustra estas modificaciones.

Ejemplo 9.3.1 Condición de zurdo

En el Ejemplo 9.2.6 consideramos un estudio sobre estudiantes universitarios ingleses y escoceses en el que se encontró que 40 de 400 estudiantes varones eran zurdos. Construyamos un intervalo de confianza del 90 % para la proporción, p , de individuos zurdos de la población⁶.

La estimación muestral de la proporción es

$$\tilde{p} = \frac{40 + 0,5(1,645^2)}{400 + 1,645^2} = \frac{40 + 1,35}{400 + 2,7} \approx 0,103$$

y el ET es

$$\sqrt{\frac{0,103(0,897)}{402,7}} = 0,015$$

Un intervalo de confianza del 90 % para p es

$$0,103 \pm (1,645)(0,015)$$

o

$$0,078 < p < 0,128$$

Por tanto, tenemos una confianza del 90 % en que entre el 7,8 % y el 12,8 % de la población muestreada sea zurda. ■

Ejercicios 9.3.1-9.3.4

9.3.1 En una muestra de 848 niños de edades entre 3 y 5 años se encontró que el 3,7 % de ellos tenían deficiencia de hierro¹⁴. Utilice estos datos para construir un intervalo de confianza del 90 % para la proporción de todos los niños con edades entre 3 y 5 años que tienen deficiencia de hierro.

9.3.2 Unos investigadores examinaron pacientes que llevaban marcapasos para ver si el uso de un teléfono móvil interfería con el funcionamiento del marcapasos. Se realizaron 959 pruebas con un tipo de teléfono móvil. Se descubrió interferencia con el marcapasos (detectada mediante monitorización del electrocardiograma) en el 15,7 % de esas pruebas¹⁵.

(a) Utilice estos datos para construir un intervalo de confianza del 90 % apropiado.

(b) ¿A qué cantidad corresponde el intervalo de confianza calculado en el apartado (a)? Responda en el contexto del ejercicio.

9.3.3 Se han encontrado mutaciones genéticas en pacientes con distrofia muscular. En un estudio se descubrió que había defectos en la codificación genética de proteínas de sarcoglicano en 23 de 180 pacientes con distrofia muscular de las extremidades¹⁶. Utilice estos datos para construir un intervalo de confianza del 99 % para la correspondiente proporción poblacional.

9.3.4 En un estudio ecológico sobre el Junco de Carolina se capturaron 53 pájaros de una cierta población. De ellos, 40 eran machos¹⁷. Utilice estos datos para construir un intervalo de confianza del 90 % para la proporción de pájaros macho en la población de Juncos de Carolina.

9.4 Inferencia sobre proporciones: el test chi-cuadrado de bondad de ajuste

En la Sección 9.2 describimos métodos para construir intervalos de confianza cuando la variable observada es categórica. Centraremos ahora nuestra atención en los contrastes de hipótesis para datos categóricos. Empezaremos considerando el análisis de una única muestra de datos categóricos. Supondremos que los datos se pueden considerar como una muestra aleatoria de alguna población y contrastaremos una hipótesis nula, H_0 , que especifica las proporciones poblacionales, o probabilidades, de las diversas categorías. He aquí un ejemplo.

Ejemplo 9.4.1 Hábitat de ciervos y fuego

¿Afecta el fuego al comportamiento de los ciervos? Seis meses después de que un incendio quemara 730 acres de hábitat homogéneo de ciervos, unos investigadores estudiaron una parcela de 3.000 acres que rodeaba el área quemada, que se dividió en cuatro regiones: región cercana al corazón del incendio (1) interior del los límites del incendio (2), exterior del límite del incendio (3) y la región exterior al área quemada (4). Véanse la Figura 9.4.1 y la Tabla 9.4.1¹⁸. La hipótesis nula es que los ciervos no muestran preferencia por un tipo particular de hábitat quemado/no quemado, es decir, que están distribuidos aleatoriamente sobre los 3.000 acres. La hipótesis alternativa es que los ciervos muestran una preferencia por alguna de las regiones, es decir, que no están aleatoriamente distribuidos sobre los 3.000 acres.

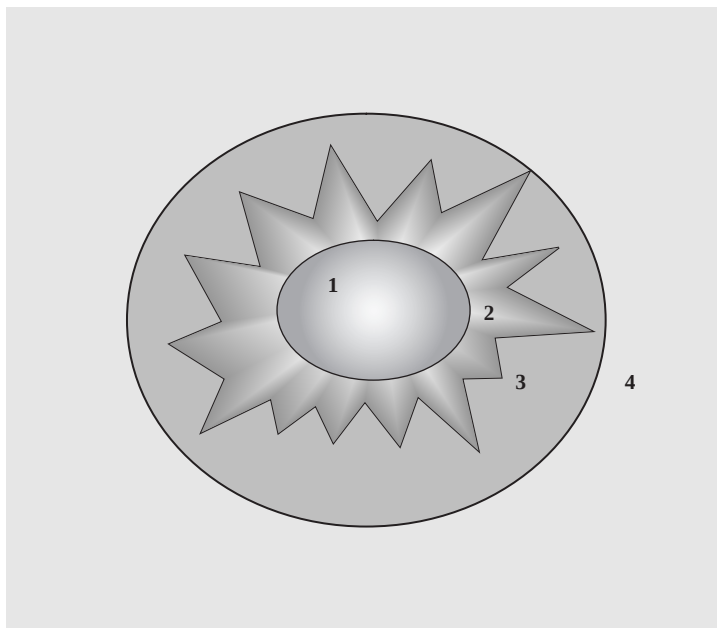


Figura 9.4.1 Esquema de la parcela de 3.000 acres con una zona interior quemada de 730 acres (no está a escala).

Tabla 9.4.1 Distribución de los ciervos		
Región	Acres	Proporción
1. Corazón zona quemada	520	0,173
2. Límite interior zona quemada	210	0,070
3. Límite exterior zona quemada	240	0,080
4. Exterior no quemado	2.030	0,677
	3.000	1.000

Bajo la hipótesis nula, si los ciervos estuvieran distribuidos aleatoriamente sobre los 3.000 acres, entonces deberíamos esperar que el número de ciervos en las regiones fuera proporcional al tamaño de dichas regiones. Expresando numéricamente la hipótesis nula, tenemos las siguientes probabilidades de ver ciervos:

$$H_0: \Pr\{\text{corazón zona quemada}\} = \frac{520}{3.000} = 0,173$$

$$\Pr\{\text{límite interior zona quemada}\} = \frac{210}{3.000} = 0,070$$

$$\Pr\{\text{límite exterior zona quemada}\} = \frac{240}{3.000} = 0,080$$

$$\Pr\{\text{exterior no quemado}\} = \frac{2.030}{3.000} = 0,677$$

Como la hipótesis alternativa no es específica (solo afirma que los ciervos prefieren algunas regiones sobre otras, pero no indica la naturaleza de la preferencia), no hay una forma simbólica simple de expresar dicha hipótesis alternativa. Por tanto, no emplearemos en general una representación simbólica. Si tuviéramos que representar simbólicamente la alternativa escribiríamos:

$$H_A: \Pr\{\text{corazón zona quemada}\} \neq 0,173, \text{ y/o } \Pr\{\text{límite zona quemada}\} \neq 0,070, \text{ y/o } \Pr\{\text{límite exterior zona quemada}\} \neq 0,080, \text{ y/o } \Pr\{\text{exterior no quemado}\} \neq 0,677$$

Dada una muestra aleatoria de observaciones categóricas, ¿cómo podemos juzgar si proporcionan evidencia contra la hipótesis nula H_0 que especifica las probabilidades de las categorías? Existen dos planteamientos complementarios sobre esta cuestión. El primero considera un examen de las frecuencias relativas observadas de cada categoría, mientras que el segundo examina las frecuencias directamente. Considerando el primer método, las frecuencias relativas observadas sirven como estimaciones de las probabilidades de las categorías. La siguiente notación para las frecuencias relativas será de utilidad: cuando una probabilidad $\Pr\{E\}$ se estime a partir de datos observados, se indicará mediante (« $\wedge$ »). Por tanto,

$$\hat{\Pr}\{E\} = \text{probabilidad estimada del suceso } E$$

Ejemplo 9.4.2 Hábitat de ciervos y fuego

Los investigadores del Ejemplo 9.4.1 observaron 75 ciervos en la parcela de 3.000 acres. Dos estaban en la región interior del incendio (Región 1), 12 en el interior del límite del incendio (Región 2), 18 en el exterior del límite del incendio (Región 3) y 43 fuera del área quemada (Región 4).

Los datos se muestran en la Figura 9.4.2.

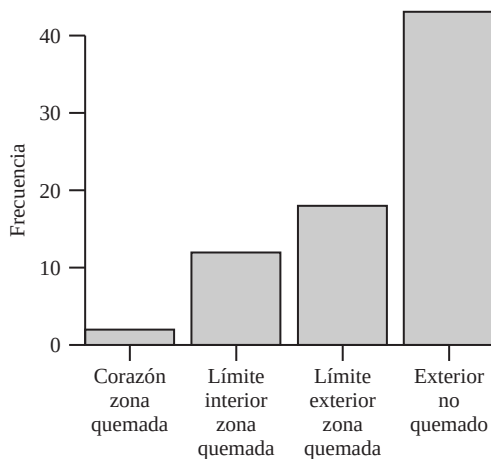


Figura 9.4.2 Diagrama de barras de los datos de distribución de ciervos

Las probabilidades estimadas de las categorías son

$$\hat{\Pr}\{\text{corazón zona quemada}\} = \frac{2}{75} = 0,027$$

$$\hat{\Pr}\{\text{límite interior zona quemada}\} = \frac{12}{75} = 0,160$$

$$\hat{\Pr}\{\text{límite exterior zona quemada}\} = \frac{18}{75} = 0,240$$

$$\hat{\Pr}\{\text{exterior no quemado}\} = \frac{43}{75} = 0,573$$

Esas probabilidades estimadas se diferencian bastante de las del modelo especificado por H_0 . La Figura 9.4.3 muestra los diagramas de barras apiladas de las distribuciones observadas y de la hipótesis nula. ■

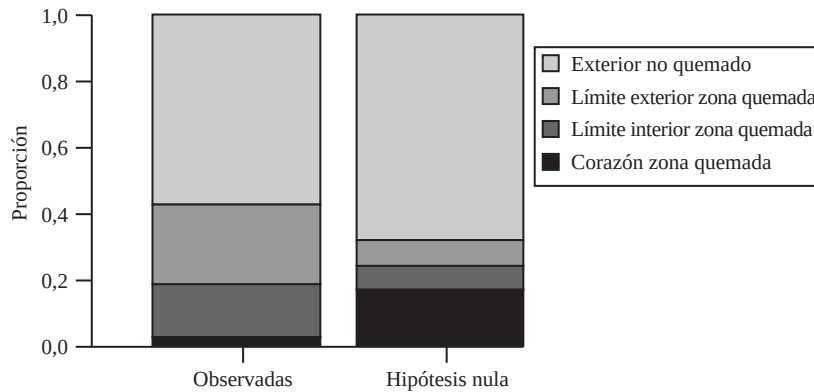


Figura 9.4.3 Diagrama de cajas apiladas de las proporciones de ciervos

El estadístico chi-cuadrado

El segundo planteamiento, que considera las frecuencias reales, es utilizar un test estadístico, denominado **test de bondad de ajuste**, para evaluar la compatibilidad de los datos con H_0 . El test de bondad de ajuste más ampliamente utilizado es el **test chi-cuadrado** o test χ^2 (χ es la letra griega «chi»).

El cálculo del estadístico de contraste chi-cuadrado se realiza en función de las frecuencias absolutas, y no relativas, de las categorías. Para cada nivel de categoría, i , sea o_i la **frecuencia observada** de dicha categoría y sea e_i la **frecuencia esperada** (es decir, la frecuencia que habría que esperar de acuerdo con H_0). Los e_i se calculan multiplicando por n cada probabilidad especificada por H_0 , como se muestra en el Ejemplo 9.4.3.

Ejemplo 9.4.3 Hábitat de ciervos y fuego

Considere la hipótesis nula especificada en el Ejemplo 9.4.1 y los datos del Ejemplo 9.4.2. Si la hipótesis nula es cierta, entonces podemos esperar que el 17,3 % de los 75 ciervos estén en el corazón de la región quemada; 17,3 % de 75 es 13,0:

$$\text{Corazón zona quemada: } e_1 = (0,173)(75) = 13,00$$

Las correspondientes frecuencias esperadas para las otras regiones son:

$$\text{Límite interior zona quemada: } e_2 = (0,070)(75) = 5,25$$

$$\text{Límite exterior zona quemada: } e_3 = (0,080)(75) = 6,00$$

$$\text{Exterior no quemado: } e_4 = (0,677)(75) = 50,75$$

El estadístico de contraste del test de bondad de ajuste chi-cuadrado se calcula a partir de o_i y e_i utilizando la fórmula dada en el recuadro siguiente con k igual al número de niveles de categorías. El Ejemplo 9.4.4 ilustra el cálculo del estadístico chi-cuadrado.

El estadístico chi-cuadrado

$$\chi_s^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

donde la suma es en todas las k categorías.

Ejemplo 9.4.4 Hábitat de ciervos y fuego

Las frecuencias observadas de las posiciones de los 75 ciervos son

Región	Corazón zona quemada	Límite interior zona quemada	Límite exterior zona quemada	Exterior no quemado	Total
Observadas (o_i)	2	12	18	43	75

Las frecuencias esperadas son

Región	Corazón zona quemada	Límite interior zona quemada	Límite exterior zona quemada	Exterior no quemado	Total
Esperadas (e_i)	13	5,25	6	50,75	75

Nótese que la suma de las frecuencias esperadas coincide con la suma de las frecuencias observadas (75). El estadístico χ^2 es:

$$\chi_s^2 = \frac{(2 - 13)^2}{13} + \frac{(12 - 5,25)^2}{5,25} + \frac{(18 - 6)^2}{6} + \frac{(43 - 50,75)^2}{50,75} = 43,2$$

Nota sobre el cálculo: Al calcular el estadístico χ^2 los valores de o_i deben ser frecuencias *absolutas*, no frecuencias relativas.

La distribución χ^2

Por la forma en la que se define χ_s^2 , resulta claro que valores pequeños de χ_s^2 deberían indicar que los datos están de acuerdo con H_0 , mientras que valores grandes de χ_s^2 deberían indicar un desacuerdo. Para basar un contraste estadístico en este acuerdo o desacuerdo, necesitamos conocer cómo puede resultar afectado χ_s^2 por la variación del muestreo.

Consideremos la distribución nula de χ_s^2 (es decir, la distribución muestral que sigue χ_s^2 si H_0 es cierta). Se puede demostrar (utilizando los métodos de la estadística matemática) que, si el tamaño de la muestra es suficientemente grande, entonces la distribución nula de χ_s^2 se puede aproximar por la distribución conocida como **distribución χ^2** . La forma de una distribución χ^2 depende de un parámetro denominado «grados de libertad» (gl). La Figura 9.4.4 muestra la distribución χ^2 con $gl = 5$.

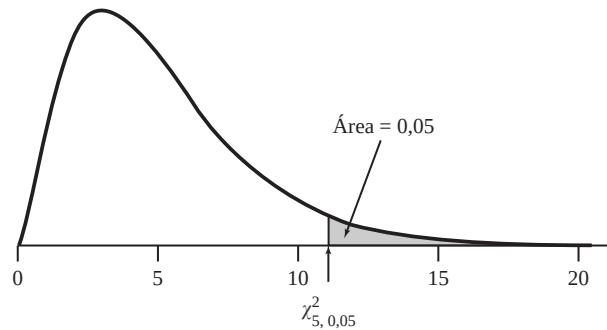


Figura 9.4.4 La distribución χ^2 con $gl = 5$

La Tabla 9 (al final del libro) contiene los valores críticos de la distribución χ^2 . Por ejemplo, para $gl = 5$, el valor crítico del 5 % es $\chi_{5, 0,05}^2 = 11,07$. Este valor crítico corresponde a un área de 0,05 en la cola superior de la distribución χ^2 , como se muestra en la Figura 9.4.4.

El test de bondad de ajuste

Para el test chi-cuadrado de bondad de ajuste que hemos presentado, la distribución nula de χ_s^2 es aproximadamente una distribución χ^2 con*

$$gl = k - 1, \text{ siendo } k \text{ el número de categorías}$$

Por ejemplo, para la situación presentada en el Ejemplo 9.4.4 hay cuatro categorías, por lo que $k = 4$. La hipótesis nula especifica las probabilidades para cada una de las cuatro categorías. Sin embargo, una vez las tres primeras probabilidades se han especificado, la última está determinada, ya que las cuatro probabilidades deben sumar 1. Hay cuatro categorías, pero solo tres de ellas son «libres». La última está restringida por las tres primeras.

El contraste de H_0 se realiza utilizando los valores críticos de la Tabla 9, como se ilustra en el siguiente ejemplo.

Ejemplo 9.4.5 Hábitat de ciervos y fuego

Para los datos sobre el hábitat de ciervos del Ejemplo 9.4.4, el estadístico chi-cuadrado observado fue $\chi_s^2 = 43,2$. Como hay cuatro categorías, los grados de libertad de la distribución nula se calculan como

$$gl = 4 - 1 = 3$$

A partir de la Tabla 9 con $gl = 3$ obtenemos que $\chi_{3, 0,0001}^2 = 21,11$. Como $\chi_s^2 = 43,2$ es mayor que 21,11, el área bajo la cola superior por encima de 43,2 es menor que 0,0001. Por tanto, el P valor es menor que 0,0001 y tenemos una evidencia fuerte en contra de H_0 y a favor de la hipótesis alternativa de que los ciervos muestran preferencia por algunas áreas sobre otras. Tras comparar las frecuencias observadas y esperadas (o, de forma equivalente, las probabilidades de la hipótesis nula y las estimadas), observamos que los ciervos se desplazan desde la regiones quemada y no quemada (1) y (4) hacia la regiones cerca de la frontera (2) y (3) (donde es probable que exista nuevo crecimiento de vegetación bajo el cobijo de la antigua). ■

El test chi-cuadrado se puede utilizar con cualquier número de categorías. En el Ejemplo 9.4.6 el test se aplica a una variable con seis categorías.

Ejemplo 9.4.6 Semillas de lino

Unos investigadores estudiaron un tipo mutante de semilla de lino que esperaban produjera aceite para su uso en la fabricación de margarina y manteca. La cantidad de ácido palmítico en la semilla de lino era un factor importante en esta investigación. Un factor relacionado era si la semilla era de color marrón o multicolor. Las semillas se clasificaron en seis combinaciones de ácido palmítico y color, como se muestra en la Tabla 9.4.2¹⁹. De acuerdo a las

Tabla 9.4.2 Distribución de las semillas de lino			
Color	Nivel de ácido	Observadas (o_i)	Esperadas (e_i)
Marrón	Bajo	15	13,5
Marrón	Intermedio	26	27
Marrón	Alto	15	13,5
Multicolor	Bajo	0	4,5
Multicolor	Intermedio	8	9
Multicolor	Alto	8	4,5
Total		72	72

* El test chi-cuadrado se puede extender a situaciones más generales en las que los parámetros se estiman de los datos antes de calcular las frecuencias esperadas. En general, los grados de libertad del test son (número de categorías) – (número de parámetros estimados) – 1. Estamos considerando únicamente el caso en el que no hay que estimar parámetros de los datos.

hipótesis de un modelo genético (mendeliano), las seis combinaciones deberían ocurrir en una proporción 3:6:3:1:2:1. Es decir, la combinación de color marrón y bajo nivel de ácido debería ocurrir una probabilidad de 3/16. La combinación de color marrón y nivel de ácido intermedio debería ocurrir con una probabilidad de 6/16, y así sucesivamente. La hipótesis nula es que el modelo es correcto. La hipótesis alternativa es que el modelo es incorrecto. El estadístico χ^2 es

$$\chi_s^2 = \frac{(15 - 13,5)^2}{13,5} + \frac{(26 - 27)^2}{27} + \frac{(15 - 13,5)^2}{13,5} + \frac{(0 - 4,5)^2}{4,5} + \frac{(8 - 9)^2}{9} + \frac{(8 - 4,5)^2}{4,5} = 7,71$$

El test χ^2 tiene $6 - 1 = 5$ grados de libertad. Observando la Tabla 9 con $gl = 5$, encontramos que $\chi_{5, 0,20}^2 = 7,29$ y $\chi_{5, 0,10}^2 = 9,24$. Por tanto, el P valor está acotado: $0,10 < P \text{ valor} < 0,20$. Si el nivel de α elegido para el contraste es 0,10 o menor, entonces el P valor es mayor que α y no deberíamos rechazar H_0 . Concluimos que no hay evidencia significativa de que los datos sean inconsistentes con el modelo mendeliano. (Nótese que no hemos demostrado necesariamente que el modelo mendeliano sea correcto, solo que no podemos rechazar este modelo). ■

Nótese que los valores críticos del test chi-cuadrado no dependen del tamaño de la muestra, n . Sin embargo, el procedimiento de contraste *está* afectado por n , a través del valor del estadístico chi-cuadrado. Si cambiamos el tamaño de la muestra manteniendo fijo el porcentaje de la composición, entonces χ_s^2 varía directamente con el tamaño de la muestra, n . Por ejemplo, imaginemos que se añade una réplica de la muestra a la propia muestra. Entonces, la muestra ampliada tendría dos veces más observaciones que la original, pero se mantendrían las mismas proporciones relativas. El valor de cada o_i sería el doble, el valor de cada e_i sería el doble y, por tanto, el valor de χ^2 sería también el doble [porque en cada término de χ_s^2 el numerador $(o_i - e_i)^2$ estaría multiplicado por 4, y el denominador e_i estaría multiplicado por 2]. Es decir, el valor de χ_s^2 aumentaría en un factor de 2, ¡a pesar del hecho de que el patrón de los datos no ha cambiado! De esta forma, un tamaño de la muestra mayor aumenta cualquier discrepancia entre lo que se observa y lo que se espera bajo la hipótesis nula.

Hipótesis compuesta y direccionalidad

Examinemos con mayor detalle la hipótesis nula del test de bondad de ajuste. En una comparación de dos muestras como un test t , la hipótesis nula contiene exactamente una afirmación (por ejemplo, que las dos medias poblacionales son iguales). Por el contrario, una hipótesis nula de un test de bondad de ajuste puede contener más de una afirmación. Una hipótesis nula de ese tipo se denomina **hipótesis nula compuesta**. Se presenta a continuación un ejemplo.

Ejemplo 9.4.7 Hábitat de ciervos y fuego

La hipótesis nula del Ejemplo 9.4.1 es

$$H_0: \Pr\{\text{corazón zona quemada}\} = 0,173, \Pr\{\text{límite interior zona quemada}\} = 0,070,$$

$$\Pr\{\text{límite exterior zona quemada}\} = 0,080, \Pr\{\text{exterior no quemado}\} = 0,677$$

Se trata de una hipótesis compuesta porque realiza tres afirmaciones independientes, concretamente

$$\Pr\{\text{corazón zona quemada}\} = 0,173, \Pr\{\text{límite interior zona quemada}\} = 0,070, \text{ y}$$

$$\Pr\{\text{límite exterior zona quemada}\} = 0,080$$

Nótese que la cuarta afirmación ($\Pr\{\text{exterior no quemado}\} = 0,677$) no es una afirmación independiente porque se deduce de las otras tres. ■

Cuando la hipótesis nula compuesta, el test chi-cuadrado tiene dos características especiales. Primero, la hipótesis alternativa es necesariamente no direccional. Segundo, si H_0 se rechaza, el contraste no conduce a una conclusión direccional. (Sin embargo, si H_0 se rechaza, entonces un examen de las proporciones observadas muestra algunas veces un patrón interesante en la forma de apartarse de H_0 , como en el Ejemplo 9.4.5).

Como H_0 es compuesta, el test chi-cuadrado es de naturaleza no direccional (quizá «omnidireccional» sería un término mejor), porque el estadístico chi-cuadrado mide desviaciones con respecto a H_0 en todas direcciones. Existen métodos estadísticos que proporcionan conclusiones direccionales y que pueden manejar alternativas direccionales, pero dichos métodos están más allá del alcance de este libro.

Variables dicotómicas

Si la variable categórica analizada por un test de bondad de ajuste es dicotómica, entonces la hipótesis nula no es compuesta, y las alternativas y conclusiones direccionales no plantean ninguna dificultad particular*.

Conclusión direccional El siguiente ejemplo ilustra la conclusión direccional.

Ejemplo 9.4.8 Hábitat de ciervos, fuego y dos regiones

Supongamos que los datos de hábitat de ciervos del Ejemplo 9.4.1 se hubieran presentado como pertenecientes solamente a dos regiones, A y B, donde la región A es el área en la frontera del incendio, que combina la regiones (2) y (3), y la región B es el resto de la parcela, es decir, la regiones (1) y (4). Se han observado 30 ciervos en la región A y 45 ciervos en la región B. ¿Representa esto una evidencia de que los ciervos prefieren una región sobre la otra?

Una hipótesis nula apropiada es

$$H_0: \Pr\{\text{región A}\} = \frac{450}{3.000} = 0,15, \Pr\{\text{región B}\} = \frac{2.550}{3.000} = 0,85$$

Esta hipótesis no es compuesta porque contiene solo una afirmación independiente. (Nótese que la segunda afirmación, $\Pr\{\text{región B}\} = 0,85$, es redundante; se deduce de la primera).

Contrastemos H_0 frente a la alternativa no direccional

$$H_A: \Pr\{\text{región A}\} \neq 0,15$$

Las frecuencias observadas y esperadas se muestran en la Tabla 9.4.3.

Tabla 9.4.3 Datos de hábitat de ciervos para dos regiones			
	A	B	Total
Observadas	30	45	75
Esperadas	11,25	63,75	75

Con los datos se calcula $\chi_s^2 = 36,8$. Utilizando la Tabla 9 obtenemos que $P < 0,0001$. Incluso con $\alpha = 0,0001$ rechazaríamos H_0 y obtendríamos que existe suficiente evidencia para concluir que la población de ciervos prefiere una región sobre la otra. Comparando los números observados y esperados, podemos ver que prefieren la región A sobre la región B. ■

Para recapitular, la conclusión direccional del Ejemplo 9.4.8 es legítima porque sabemos que si H_0 es falsa, entonces necesariamente $\Pr\{\text{región A}\} < 0,15$ o $\Pr\{\text{región A}\} > 0,15$. Por el contrario, en el Ejemplo 9.4.7, H_0 puede ser falsa pero $\Pr\{\text{exterior no quemado}\}$ podría ser todavía igual a 0,677. El análisis chi-cuadrado no determina cuáles de las probabilidades no son las especificadas por H_0 .

Alternativa direccional Un test chi-cuadrado de bondad de ajuste contra una alternativa direccional (cuando la variable observada es dicotomía) utiliza el procedimiento familiar en dos pasos:

Paso 1. Comprobar la direccionalidad (ver si los datos se desvían con respecto a H_0 en la dirección especificada por H_A).

- (a) Si no es así, el P valor es mayor que 0,50.
- (b) Si es así, continuar con el paso 2.

* Cuando los datos son dicotómicos, hay una alternativa al test de bondad de ajuste que se conoce como test Z de una sola proporción. Los cálculos utilizados en el test Z parecen bastante diferentes de los del test de bondad de ajuste, pero de hecho los dos tests son matemáticamente equivalentes. Sin embargo, a diferencia del test de bondad de ajuste, donde se puede manejar cualquier número de categorías, el test Z solo se puede utilizar cuando los datos están limitados a dos categorías. Por tanto, no lo presentaremos aquí.

Paso 2. El P valor es la mitad del que sería si H_A fuera no direccional.

El siguiente ejemplo ilustra el procedimiento.

Ejemplo 9.4.9 Festival de la Luna de la Cosecha

¿Pueden las personas cercanas a la muerte posponerla hasta después de haber transcurrido una ocasión simbólicamente significativa? Unos investigadores estudiaron las muertes por causas naturales entre mujeres ancianas chinas (por encima de 75 años) que vivían en California. Escogieron estudiar el tiempo alrededor del Festival de la Luna de la Cosecha porque (1) la fecha de ese festival tradicional chino cambia algo de un año a otro, haciendo así menos probable que un efecto debido a la época del año se confundiera con el efecto que estaban estudiando y (2) es un festival en el cual el papel de las mujeres ancianas en la familia es muy importante.

Investigaciones previas habían sugerido que podría haber una disminución en la tasa de mortalidad de las mujeres ancianas chinas inmediatamente antes del festival, con un incremento correspondiente después. Los investigadores descubrieron que durante un periodo de varios años hubo 33 muertes en el grupo durante la semana que precedía al Festival de la Luna de la Cosecha y 70 muertes en la semana siguiente al festival²⁰. ¿Con qué fuerza apoyan estos datos la interpretación de que las personas pueden prolongar su vida hasta que transcurre un evento simbólicamente significativo?

Podemos formular las hipótesis nula y alternativa como sigue:

H_0 : Sabiendo que una mujer anciana china muere en el intervalo de una semana antes o después del Festival de la Luna de la Cosecha, es igualmente probable que muera la semana antes o la semana después.

H_A : Sabiendo que una mujer anciana china muere en el intervalo de una semana antes o después del Festival de la Luna de la Cosecha, es más probable que muera la semana después que la semana antes.

Las hipótesis anteriores se pueden expresar como

$$H_0: \Pr\{\text{morir después del festival}\} = \frac{1}{2}$$

$$H_A: \Pr\{\text{morir después del festival}\} > \frac{1}{2}$$

donde se entiende que $\Pr\{\text{morir después del festival}\}$ es la probabilidad de morir después del festival, sabiendo que la mujer muere en el intervalo de una semana antes o después del festival. Las frecuencias observadas y esperadas se muestran en la Tabla 9.4.4.

Tabla 9.4.4 Datos del Festival de la Luna de la Cosecha			
	Antes	Después	Total
Observadas	33	70	103
Esperadas	51,5	51,5	103

Observando los datos de las 103 muertes, notamos primero que los datos, de hecho, se desvían con respecto a H_0 en la dirección especificada por H_A , ya que la frecuencia relativa observada de muertes después del festival de 70/103, que es mayor que 1/2. El valor del estadístico chi-cuadrado es $\chi_s^2 = 13,3$. Utilizando la Tabla 9, podemos ver que el P valor estaría acotado entre 0,0001 y 0,001 si H_A fuera no direccional. Sin embargo, para la hipótesis alternativa direccional especificada en este contraste, el P valor se acota como $0,00005 < P \text{ valor} < 0,0005$. Concluimos por tanto que la evidencia es muy fuerte a favor de que las tasas de muertes entre las mujeres ancianas chinas aumentan después del festival*.

* Basándose en estos resultados, se podría extraer la conclusión de que habría que cancelar el festival para proteger a las mujeres ancianas chinas. Como este estudio es solo observacional, ¡no deberíamos extraer conclusiones causales!

Ejercicios 9.4.1-9.4.13

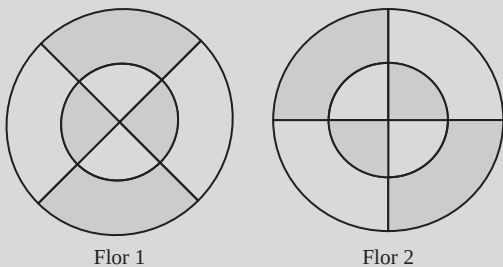
9.4.1 Un cruce entre calabacines de verano blancos y amarillos produce descendencia de los siguientes colores²¹:

Color	Blanco	Amarillo	Verde
Número de descendientes	155	40	10

¿Son estos datos consistentes con una proporción de 12:3:1 predicha por un cierto modelo genético? Utilice un test chi-cuadrado con $\alpha = 0,10$.

9.4.2 En referencia al Ejercicio 9.4.1, supongamos que la muestra tuviera la misma composición pero fuera 10 veces más grande: una descendencia de 1.550 blancos, 400 amarillos y 100 verdes. ¿Serían los datos consistentes con el modelo 12:3:1?

9.4.3 ¿Cómo reconocen las abejas a las flores? Como parte de un estudio sobre esta cuestión, unos investigadores utilizaron dos «flores» artificiales²²:



El experimento consistió en una serie de pruebas con abejas individuales. Cada prueba consistía en presentar a la abeja ambas flores y observar en qué flor se posaba primero. (La flor 1 estaba algunas veces a la izquierda y otras a la derecha). Durante las pruebas de «entrenamiento», la flor 1 contenía una solución de sacarosa y la flor 2 no. Por tanto, se entrenó a las abejas para que prefirieran la flor 1. Durante las pruebas del test, ninguna flor contenía sacarosa. En 25 pruebas de test con una abeja concreta, la abeja eligió la flor 1 veinte veces y la flor 2 cinco veces.

Utilice un test de bondad de ajuste para evaluar la evidencia a favor de que la abeja podría recordar y distinguir los patrones de flores. Utilice una alternativa direccional y $\alpha = 0,05$.

9.4.4 En un hospital del medio oeste hubo un total de 932 nacimientos en 20 semanas consecutivas. De esos nacimientos, 216 ocurrieron en fin de semana²³. ¿Revelan estos datos una desviación no debida al azar de un patrón aleatorio de los nacimientos? (Realice un test de bondad de ajuste con dos categorías de nacimientos: día entre semana y día de fin de semana. Utilice una alternativa direccional y $\alpha = 0,05$).

9.4.5 En un experimento sobre reproducción se aparearon gallinas blancas de cresta pequeña y produjeron 190 descendientes de los tipos que se muestran en la tabla que acompaña al ejercicio²⁴. ¿Son estos datos consistentes con proporciones mendelianas esperadas de 9:3:3:1 para los cuatro tipos? Utilice un test chi-cuadrado y $\alpha = 0,10$.

Tipo	Número de descendientes
Plumas blancas, cresta pequeña	111
Plumas blancas, cresta grande	37
Plumas oscuras, cresta pequeña	34
Plumas oscuras, cresta grande	8
Total	190

9.4.6 De n bebés nacidos en una cierta ciudad, el 51 % fueron niños²⁵. Suponga que se desea contrastar la hipótesis de que la verdadera probabilidad de que el bebe sea niño es $\frac{1}{2}$.

Calcule el valor de χ_s^2 , y acote el P valor para realizar el contraste contra una alternativa no direccional, si

- (a) $n = 1.000$ (b) $n = 5.000$ (c) $n = 10.000$

9.4.7 En un experimento agrónomo se cruzaron cacahuets con semillas arrugadas con cacahuets normales. El modelo genérico considerado por los agrónomos predecía que la proporción de descendientes normales frente a arrugados sería de 3:1. Obtuvieron 95 descendientes normales y 54 arrugados²⁶. ¿Dan soporte los datos al modelo de hipótesis?

Realice un contraste chi-cuadrado con $\alpha = 0,05$. Utilice una alternativa no direccional.

9.4.8 Se realizó un diseño experimental basado en ajuste de camadas para probar una cierta sustancia que podría causar cáncer. Partiendo de 50 camadas de ratas, en cada una de ellas se seleccionaron tres hembras. Una de las tres, seleccionada aleatoriamente, recibió la sustancia de prueba, y las otras dos se utilizaron como controles. Durante un periodo de observación de dos años se apuntó para cada animal el instante de aparición de un tumor y/o la muerte por diversas causas. Una forma de analizar los casos es apuntar simplemente qué rata (en cada terna) desarrolla un tumor en primer lugar. Algunas ternas no proporcionan información en este sentido porque (a) ninguna de las ratas de la terna desarrolla un tumor o (b) una rata desarrolla un tumor después de que otra compañera de terna muera por otra causa. Los resultados para las 50 ternas se muestran en la tabla siguiente²⁷. Utilice un test de bondad de ajuste para evaluar la evidencia de que la sustancia causa cáncer. Utilice una alternativa direccional y $\alpha = 0,01$. Plantee su conclusión en el contexto

del ejercicio. (Sugerencia: utilice solo las 20 ternas que proporcionan información completa).

	Número de ternas
Primer tumor en una rata tratada	12
Primer tumor en una de las ratas de control	8
Ningún tumor	23
Muerte por otra causa	7
Total	50

9.4.9 En un estudio sobre la visión del color de las ardillas se utilizó un aparato que contenía tres pequeños paneles translúcidos que se podían iluminar de forma separada. Los animales se entrenaron para escoger, accionando una palanca, el panel que parecía diferente a los otros dos. (Durante las sesiones de «entrenamiento» los paneles se diferenciaban en brillo, en vez de en color). Después se probó la capacidad de los animales de discriminar entre diversos colores. En una de las series de «prueba» de un animal, uno de los paneles era rojo y los otros dos eran blancos. La situación del panel rojo se modificaba aleatoriamente de una prueba a otra. En 75 pruebas, el animal escogió correctamente 45 veces e incorrectamente 30 veces²⁸. ¿Con qué fuerza dan soporte estos datos a la interpretación de que el animal puede discriminar entre los dos colores?

- (a) Contraste la hipótesis nula de que el animal no puede discriminar el rojo del blanco. Utilice una alternativa direccional y $\alpha = 0,02$.
- (b) ¿Por qué es adecuada una alternativa direccional en este caso?

9.4.10 Los científicos han utilizado los jerbos de Mongolia en investigaciones neurológicas. Se cruzó una cierta raza de jerbos cuya descendencia presentó los siguientes colores²⁹:

Color	Negro	Marrón	Blanco
Número de descendientes	40	59	42

¿Son estos datos consistentes con una proporción de 1:2:1 predicha por cierto modelo genético? Utilice un test chi-cuadrado con $\alpha = 0,05$.

9.4.11 Se pidió a 36 hombres que tocaran con los ojos vendados la frente de tres mujeres, una de las cuales era su pareja. Las dos mujeres «señuelo» eran de la misma edad, altura y peso que la pareja del hombre. De los 36 hombres, 18 fueron capaces de reconocer correctamente a su pareja³⁰. ¿Proporcionan estos datos evidencia suficiente para concluir que los hombres lo han hecho mejor que simplemente adivinando?

Realice un contraste apropiado.

9.4.12 Unos genetistas que estudiaban el patrón de herencia de las plantas de caupí clasificaron las plantas de un experimento de acuerdo a la naturaleza de sus hojas. Los datos se presentan a continuación³¹:

Tipo	I	II	III
Número	179	44	23

Contraste la hipótesis nula de que los tres tipos aparecen con probabilidades 12/16, 3/16 y 1/16. Utilice un test chi-cuadrado con $\alpha = 0,10$.

9.4.13 En la boca de dragón (*Antirrhinum majus*), las plantas pueden tener flores rojas, rosas o blancas. De acuerdo a un cierto modelo genético mendeliano, la autopolinización de plantas con flores rosas debería producir una descendencia con flores rojas, rosas y blancas en la proporción 1:2:1. Un genetista autopolinizó bocas de dragón de flores rosas y produjo una descendencia de 234 individuos con los siguientes colores³²:

Tipo	Rojo	Rosa	Blanco
Número	54	122	58

Contraste la hipótesis nula de que los tres colores aparecen con probabilidades 1/4, 1/2, y 1/4. Utilice un test chi-cuadrado con $\alpha = 0,10$.

9.5 Perspectiva y resumen

En este capítulo hemos presentado la inferencia para datos divididos en categorías, incluyendo los intervalos de confianza y los contrastes de hipótesis. Los procedimientos desarrollados, que se resumen seguidamente, se pueden aplicar si (1) los datos se pueden considerar como una muestra aleatoria de una población grande y (2) las observaciones son independientes.

Resumen de los métodos de inferencia para datos divididos en categorías**Intervalo de confianza del 95 % para p**

$$\tilde{p} \pm 1,96 \times ET_{\tilde{p}}$$

siendo

$$\tilde{p} = \frac{y + 2}{n + 4}$$

y

$$ET_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}$$

Intervalo de confianza general para p

$$\tilde{p} \pm z_{\alpha/2} \times ET_{\tilde{p}}$$

siendo

$$\tilde{p} = \frac{y + 0,5(z_{\alpha/2}^2)}{n + z_{\alpha/2}^2}$$

$$ET_{\tilde{p}} = \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + z_{\alpha/2}^2}}$$

Test de bondad de ajuste*Datos:* o_i = frecuencia observada de la categoría i *Hipótesis nula:* H_0 especifica la probabilidad de cada categoría*.*Cálculo de las frecuencias esperadas:* $e_i = n \times$ Probabilidad especificada en H_0 para la categoría i *Estadístico de contraste:*

$$\chi_s^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

*Distribución nula (aproximada):*Distribución χ^2 con $gl = k - 1$ siendo k = el número de categorías.La aproximación es adecuada si $e_i \geq 5$ en cada categoría.

* Se puede utilizar una forma ligeramente modificada del test de bondad de ajuste para contrastar una hipótesis que simplemente restringe las probabilidades en vez de especificarlas completamente. Un ejemplo sería el contraste del ajuste de una distribución binomial a unos datos (véase la Sección opcional 3.9). Los detalles de este contraste están fuera del alcance de este libro.

Ejercicios suplementarios 9.S.1-9.S.21

9.S.1 En una cierta población, el 83 % de las personas tiene sangre tipo Rh positivo³³. Suponga que se toma una muestra aleatoria de $n = 10$ personas de la población y sea $\tilde{P}$ la proporción ajustada de Wilson de las personas con Rh positivo en la muestra. Calcule

- (a) $\Pr\{\tilde{P} = 0,714\}$ (b) $\Pr\{\tilde{P} = 0,786\}$

9.S.2 En una población de gusanos platelmintos (*Planaria*) que viven en una cierta laguna, uno de cada cinco individuos es adulto y cuatro son jóvenes³⁴. Una ecóloga planea contar los adultos en una muestra aleatoria de 16 gusanos de la laguna. Utilizará después $\tilde{P}$, la proporción muestral ajustada de Wilson de adultos en la muestra, como estimador de p , la proporción de adultos en la población de la laguna. Calcule

- (a) $\Pr\{\tilde{P} = p\}$ (b) $\Pr\{p - 0,05 \leq \tilde{P} \leq p + 0,05\}$

9.S.3 En un estudio sobre los efectos del entorno sobre la reproducción se capturaron 123 ciervos de cola blanca hembra adultos y se encontró que 97 estaban embarazadas³⁵. Construya un intervalo de confianza del 95 % para la proporción de hembras embarazadas en la población de ciervos.

9.S.4 En referencia al Ejercicio 9.S.3, ¿cuál de las condiciones para la validez del intervalo de confianza se podría haber violado en este estudio?

9.S.5 En una muestra de 32 bebés criados con lactancia materna se encontró que 2 de ellos habían desarrollado deficiencia de hierro a la edad de 5,5 meses³⁶.

- (a) Utilice estos datos para construir un intervalo de confianza del 90 % apropiado.
 (b) ¿Qué condiciones son necesarias para que el intervalo de confianza del apartado (a) sea válido?
 (c) Interprete el intervalo de confianza obtenido en el apartado (a) en el contexto de este ejercicio. Es decir, ¿qué indican los números del intervalo de confianza sobre la deficiencia en hierro en bebés criados con lactancia materna?

9.S.6 Una cierta bodega de California produce 720.000 botellas de vino al año. Suponga que se desea estimar la proporción de esas botellas que tienen el corcho defectuoso (es decir, que el vino se ha estropeado debido a un fallo del corcho). Suponga que el 4 % de todas las botellas tienen un fallo en el corcho. Utilizando esta estimación inicial de p , ¿cuántas botellas de vino sería necesario incluir en una muestra aleatoria si se deseara que el error típico del estimador fuera menor o igual que un punto porcentual?³⁷

9.S.7 En referencia al Ejercicio 9.S.6, suponga que no confía en que la probabilidad de fallo del corcho del 4 % de los vinos sea en general una estimación inicial útil para una bodega en particular.

(a) Suponga que, basándose en los datos de años anteriores de esta bodega, aproximadamente el 10 % de los vinos tienen el corcho defectuoso. ¿Cuántas botellas habría que incluir en una muestra aleatoria si se desea que el error típico del estimador sea menor o igual que 1 punto porcentual?

(b) ¿Cuántas botellas habría que incluir en una muestra aleatoria si se desea que el error típico del estimador sea menor o igual que un punto porcentual, independientemente del valor de p ?

9.S.8 Cuando se agrupan ratones macho, uno de ellos se hace finalmente dominante sobre los demás. Para ver cómo puede afectar una infección parasitaria en la competencia por el dominio, se alojaron ratones macho en grupos de tres ratones por jaula. Dos de los ratones de cada jaula recibieron una dosis suave del gusano parásito *H. polygyrus*. Dos semanas después se usaron criterios tales como la ausencia de heridas en la cola para identificar al ratón dominante de cada jaula. Se encontró que el ratón no infectado era dominante en 15 de las 30 jaulas³⁸. ¿Representa esto evidencia de que la infección parasitaria tiende a inhibir el desarrollo del comportamiento dominante? Utilice un test de bondad de ajuste contra una alternativa direccional. Use $\alpha = 0,05$. (Sugerencia: la unidad observacional en este experimento no es un ratón individual, sino la jaula de tres ratones).

9.S.9 ¿Son los ratones diestros o zurdos? En un estudio sobre esta cuestión, se comprobó la pata preferida de 320 ratones una cepa altamente endogámica, observando qué pata delantera (derecha o izquierda) utilizaban para coger comida de un tubo estrecho. Cada animal realizó la prueba 50 veces, con un total de $320 \times 50 = 16.000$ observaciones. Los resultados fueron los siguientes³⁹:

	Derecha	Izquierda
Número de observaciones	7.871	8.129

Suponga que se asigna a cada categoría una frecuencia esperada de 8.000 y se realiza un test de bondad de ajuste. Se obtiene que $\chi_s^2 = 4,16$, por lo que con $\alpha = 0,05$ se rechazaría la hipótesis de una proporción de 1:1 y se obtendría evidencia suficiente para concluir que los ratones de esta cepa están (ligeramente) sesgados hacia el uso de la pata izquierda. Este análisis contiene un error fatal. ¿Cuál es?

9.S.10 Como parte del estudio del patrón de herencia de plantas de caupí, unos genetistas clasificaron las plantas de un experimento en función de si tenían una hoja o tres. Los datos fueron⁴⁰:

Número de hojas	1	3
Número de plantas	74	61

Contraste la hipótesis nula de que los dos tipos de plantas aparecen con igual probabilidad. Utilice una alternativa no direccional y $\alpha = 0,05$.

9.S.11 La gente que recolecta setas silvestres come algunas veces la seta tóxica «sombrero mortal», *Amanita phalloides*. En una revisión de 205 casos de envenenamiento por esta seta en Europa entre 1971 y 1980, unos investigadores encontraron que 45 de las víctimas murieron⁴¹. Realice un contraste para comparar esta mortalidad con la mortalidad del 30 % que se había obtenido antes de 1970. Utilice la hipótesis alternativa de que la mortalidad ha disminuido y use $\alpha = 0,05$.

9.S.12 La aparición de glándulas de pigmento en el estado de semillero de plantas de algodón se controla genéticamente. De acuerdo con una teoría del mecanismo de control, la proporción poblacional de plantas con glándulas a plantas sin glándulas resultado de un cierto cruce debería ser de 11:5. De acuerdo con otra teoría debería ser de 13:3. En un experimento, el cruce produjo 89 plantas con glándulas y 36 sin glándulas⁴². Utilice un test de bondad de ajuste (con $\alpha = 0,10$) para determinar si los datos son consistentes con

- (a) La teoría de 11:5. (b) La teoría de 13:3.

9.S.13 (Continuación del Ejercicio 9.S.12)

- (a) Si las proporciones de 11:5 y 13:3 son solo dos teorías razonables a considerar, ¿habría evidencia convincente de que la teoría seleccionada en el Ejercicio 9.S.12 es la teoría correcta? Explique su respuesta.
- (b) Si existieran otras proporciones posibles que no han sido consideradas, ¿habría evidencia convincente de que la teoría seleccionada en el Ejercicio 9.S.12 es la teoría correcta? Explique su respuesta.

9.S.14 Cuando huye de un depredador, el pez pequeño de agua dulce *Fundulus notti* se dirige frecuentemente hacia la ribera y salta sobre la orilla. En un estudio sobre la orientación espacial de este pez, se capturaron ejemplares en varias localizaciones y posteriormente se hicieron pruebas en un estanque para ver qué dirección escogían cuando se soltaban. ¿Nadarían en una dirección tal que, en su lugar de captura, les hubiera conducido a la orilla? Se presentan a continuación las direcciones escogidas ($\pm 45^\circ$) por 50 peces que realizaron la prueba en condiciones de cielo nublado⁴³.

Hacia la orilla	18
Alejándose de la orilla	12
Siguiendo la orilla hacia la derecha	13
Siguiendo la orilla hacia la izquierda	7

Utilice el test chi-cuadrado con $\alpha = 0,05$ para contrastar la hipótesis de que la elección de dirección bajo un cielo nublado es aleatoria:

- (a) Utilizando las cuatro categorías indicadas en la tabla.

- (b) Reduciendo las categorías a dos: «hacia la orilla» y «alejándose o a lo largo de la orilla» y empleando una H_A direccional.

(Nota: Aunque el test chi-cuadrado es válido en este caso, debería tenerse en cuenta que existen tests más potentes para el análisis de datos de orientación)⁴⁴.

9.S.15 Utilizando los datos de peso del córtex del Ejercicio 8.4.4:

- (a) Utilice un test de bondad de ajuste para contrastar la hipótesis de que el control del entorno no tiene ningún efecto. Como en el Ejercicio 8.4.4, utilice una alternativa direccional y $\alpha = 0,05$. (Este ejercicio muestra cómo, modificando el punto de vista, el test de signos se puede reinterpretar como un test de bondad de ajuste. Por supuesto, el test chi-cuadrado de bondad de ajuste descrito en este capítulo solo se puede usar cuando el número de observaciones es suficientemente grande).
- (b) ¿Es el número de observaciones suficientemente grande para que el test del apartado (a) sea válido?

9.S.16 Una bióloga deseaba saber si el gorgojo de la caupí tenía preferencia por un tipo de semilla sobre otras a la hora de depositar sus huevos. Puso cantidades iguales de cuatro tipos de semillas en un tarro y añadió gorgojos adultos. Tras unos pocos días, observó los siguientes datos⁴⁵:

Tipo de semilla	Número de huevos
Pinta	167
Caupí	176
Judías Navy	174
Judías del norte	194

¿Proporcionan estos datos evidencia de una preferencia por algún tipo de semilla sobre las otras? Es decir, ¿son los datos consistentes con la afirmación de que los huevos se distribuyen aleatoriamente entre los cuatro tipos de semilla?

9.S.17 En un experimento se cruzaron dos tipos de calabacines. De acuerdo con un modelo genérico, 1/2 de las plantas resultantes deberían tener tallos y frutos oscuros, 1/4 deberían tener tallos y frutos claros, y 1/4 deberían tener tallo claro y fruto liso. Los datos reales para las tres categorías fueron 220, 129 y 105⁴⁶. ¿Refuta los datos el modelo? Realice un test chi-cuadrado con $\alpha = 0,10$.

9.S.18 Se pidió a 36 hombres que tocaran con los ojos vendados las manos y la espalda de tres mujeres, una de las cuales era su pareja. Las dos mujeres «señuelo» eran de la misma edad, altura y peso que las parejas correspondientes³⁰. De los 36 hombres de la prueba, 16 fueron capaces de identificar correctamente a su pareja. ¿Proporcionan los datos evidencia suficiente a favor de que los hombres pueden identificar a su pareja mejor de lo que lo harían si simple-

mente intentan adivinar? Realice un test de bondad de ajuste de los datos con $\alpha = 0,05$.

9.S.19 En un estudio sobre resistencia a un cierto virus de la soja, unos biólogos realizaron fertilización cruzada de dos cultivos de soja. Esperaban obtener una proporción de plantas resistentes frente a susceptibles de 3:1. Los datos observados fueron de 58 plantas resistentes y 26 susceptibles⁴⁷. ¿Son estos datos significativamente inconsistentes con la proporción de 3:1 esperada? Realice un test, utilizando $\alpha = 0,10$ y una alternativa no direccional.

9.S.20 Se aconsejó sobre el uso del preservativo y el riesgo de contraer una enfermedad de transmisión sexual (ETS) a un grupo de 1.438 pacientes sexualmente activos. Después de seis meses, 103 pacientes habían contraído una ETS⁴⁸. Construya un intervalo de confianza del 95 % para la probabilidad de contraer una ETS dentro de intervalo de los seis

meses posteriores a haber participado en un programa de asesoría como el utilizado en este estudio.

9.S.21 (Continuación del Ejercicio 9.S.20) Suponga que para individuos sexualmente activos que no participaron en el programa, la probabilidad de contraer una ETS en un periodo de seis meses es del 10 %.

- Utilizando el intervalo calculado en el Ejercicio 9.S.21, ¿existe evidencia convincente de que la probabilidad de ETS en seis meses es diferente de los que participaron en el programa?
- Utilizando los datos del Ejercicio 9.S.21, realice un test chi-cuadrado no direccional para determinar si la probabilidad a los seis meses es diferente en los individuos que participaron y no participaron en el programa.
- ¿Concuerdan sus respuestas a los apartados (a) y (b)? Explique su respuesta.

Notas

- White, A. S., Godard, R. D., Belling, C., Kasza, V. y Beach, R. L. (2010). Beverages obtained from soda fountain machines in the U.S. contain microorganisms, including coliform bacteria. *International Journal of Food Microbiology* **137**, 61-66.
- From the National Survey of Family Growth (2002). U.S. Dept. of Health and Human Services, *Vital and Health Statistics*, Serie 23, n.º 25. Datos tomados de la Tabla 75.
- Agresti, A. y Coull, B. A. (1998). Approximate is better than «exact» for interval estimation of binomial proportions. *The American Statistician* **52**, 119-126. Los autores demuestran que los intervalos de confianza del 95 % basados en $\tilde{p}$ son superiores a otros intervalos de confianza utilizados comúnmente. Hacen notar también que si se utiliza $\tilde{p}$, no es necesario construir tablas o emplear reglas para saber el tamaño muestral necesario para que el intervalo de confianza tenga buenas propiedades de cobertura.
- Couch, F. J., et al. (1997). *BRCA1* mutations in women attending clinics that evaluate the risk of breast cancer. *New England Journal of Medicine* **336**, 1409-1415.
- Ware, J. H. (1989). Investigating therapies of potentially great benefit: ECMO. *Statistical Science* **4**, 298-306. Los datos se presentan con gran detalle en la Sección 10.4.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97-113.
- Adaptado de McCloskey, R. V., Goren, R., Bissett, D., Bentley, J. y Tutlane, V. (1982). Cefotaxime in the treatment of infections of the skin and skin structure. *Reviews of Infectious Diseases* **4**, supp., S444-S447.
- Adaptado de Petras, M. L. (1967). Studies of natural populations of *Mus*. III. Coat color polymorphisms. *Canadian Journal of Genetic Cytology* **9**, 287-296.
- Miller, C. L., Pollock, T. M. y Clewer, A. D. F. (1974). Whooping-cough vaccination: An assessment. *The Lancet* **ii**, 510-513.
- Erskine, A. G. y Socha, W. W. (1978). *The Principles and Practices of Blood Grouping*. St. Louis: Mosby, p. 209.
- Curtis, H. (1983). *Biology*, 4.^a ed. New York: Worth, p. 908.
- Mourant, A. E., Kopec, A. C. y Domaniewska-Sobczak, K. (1976). *The Distribution of Human Blood Groups and Other Polymorphisms*, 2.^a ed. London: Oxford University Press, p. 44.
- Basado en un experimento descrito en Oellerman, C. M., Patterson, F. L. y Gallun, R. L. (1983). Inheritance of resistance in «Luso» wheat to Hessian fly. *Crop Science* **23**, 221-224.
- Cogswell, M. E., Looker, A. C., Pfeiffer, C. M., Cook, J. D., Lacher, D. A., Beard, J. L., Lynch, S. R. y Grummer-Strawn, L. M. (2009). Assessment of iron deficiency in US preschool children and nonpregnant females of childbearing age: National Health and Nutrition Examination Survey 2003-2006. *American Journal of Clinical Nutrition* **89**, 1334-1342.
- Hayes, D. L., et al. (1997). Interference with cardiac pacemakers by cellular telephones. *New England Journal of Medicine* **336**, 1473-1479. Los datos citados son para teléfonos CDMA. Aunque se advirtió interferencia en el

15,7 % de los tests, un porcentaje mucho menor causó síntomas clínicamente significativos.

16. Duggan, D. J., Gorospe, J. R., Fanin, M., Hoffman, E. P. y Angelini, C. (1997). Mutations in the sarcoglycan genes in patients with myopathy. *New England Journal of Medicine* **336**, 618-624.

17. Rabenold, K. R. y Rabenold, P. P. (1985). Variation in altitudinal migration, winter segregation, and site tenacity in two subspecies of dark-eyed juncos in the Southern Appalachians. *The Auk* **102**, 805-819.

18. Datos ficticios pero realistas basados en una comunicación personal con F. Villablanca.

19. Saeidi, G. y Rowland, G. G. (1997). The inheritance of variegated seed color and palmitic acid in flax. *Journal of Heredity* **88**, 466-468.

20. Phillips, D. P. y Smith, D. G. (1990). Postponement of death until symbolically meaningful occasions. *Journal of the American Medical Association* **263**, 1947-1951. A efectos de comparación, los autores también examinaron fallecimientos entre mujeres ancianas judías durante el mismo periodo de tiempo; no encontraron fallecimientos en exceso después del festival en este grupo de comparación.

21. Sinnott, E. W. y Durham, G. B. (1922). Inheritance in the summer squash. *Journal of Heredity* **13**, 177-186.

22. Adaptado de Gould, J. L. (1985). How bees remember flower shapes. *Science* **227**, 1492-1494. Figure copyright 1985 de la American Association for the Advancement of Science; utilizado con permiso.

23. Adaptado de datos de nacimiento en 1983 en West Lafayette, Indiana.

24. Bateson, W. y Saunders, E. R. (1902). *Reports to the Evolution Committee of the Royal Society* **1**, 1-160. El color de las plumas y la forma del peine se controlaron independientemente; plumas blancas y peine pequeño son dominantes. Los padres del experimento eran híbridos de primera generación (F_1) y, por tanto, fueron necesarios heterocigotos dobles.

25. Es un valor realista. Véase el Ejercicio 3.6.6.

26. Jakkula, L. R., Knault, D. A. y Gorbet, D. W. (1997). Inheritance of a shriveled seed trait in peanut. *Journal of Heredity* **88**, 47-51. Los datos están tomados de la Tabla 5 del artículo.

27. Adaptado de Mantel, N., Bohidar, N. R. y Ciminera, J. L. (1977). Mantel-Haenszel analyses of litter-matched time-to-response data, with modifications for recovery of inter-litter information. *Cancer Research* **37**, 3863-3868. (En el artículo se describe un análisis más potente que utiliza las ternas parcialmente informativas).

28. Adaptado de Jacobs, G. H. (1978). Spectral sensitivity and colour vision in the ground-dwelling sciurids: Results from golden mantled ground squirrels and comparisons for five species. *Animal Behaviour* **26**, 409-421. Véase también

Jacobs, G. H. (1981). *Comparative Color Vision*, Academic Press.

29. Petrij, F., van Veen, K., Mettler, M. y Bruckmann, V. (2001). A second acromelanistic allelomorph at the albino locus of the Mongolian gerbil (*Meriones unguiculatus*). *Journal of Heredity* **92**, 74-78. Los jerbos que denominamos «marrones» son denominados «Siameses» por los autores.

30. Kaitz, M. (1992). Recognition of familiar individuals by touch. *Physiology and Behavior* **52**, 565-567.

31. Fawole, I. (2001). Genetic analysis of mutations at loci controlling leaf form in cowpea (*Vigna unguiculata* [L.] Walp). *Journal of Heredity* **92**, 43-50. Estos datos provienen de la generación 1993a recogidos en la Tabla 8 del artículo. Los tipos que denominamos I, II, y III son identificados por los autores como trifoliado, orbicular trifoliado y orbicular unifoliado.

32. Baur, E., Fischer, E. y Lenz, F. (1931). *Human Heredity*, 3.^a ed. New York: Macmillan, p. 52.

33. Esto es típico en poblaciones de EE.UU. Véase, por ejemplo, Maccready, R. A. y Mannin, M. C. (1951). A typing study of one hundred and fifty thousand bloods. *Journal of Laboratory and Clinical Medicine* **37**, 634-636.

34. Situación ficticia pero realista. Véase Krebs, C. J. (1972). *Ecology: The Experimental Analysis of Distribution and Abundance*. New York: Harper and Row.

35. Cheatum, F. L. y Severinghaus, C. W. (1950). Variations in fertility of white-tailed deer related to range conditions. *Transactions of the North American Wildlife Conference* **15**, 170-189.

36. Ziegler, E. E., Nelson, S. E. y Jeter, J. M. (2009). Iron supplementation of breastfed infants from an early age. *American Journal of Clinical Nutrition* **89**, 525-532.

37. Fischer, C. y Fischer, U. (1997). Analysis of cork taint in wine and cork material at olfactory sub-threshold levels by solid phase microextraction. *Journal of Agricultural and Food Chemistry* **45**, 1995-1997.

38. Freeland, W. J. (1981). Parasitism and behavioral dominance among male mice. *Science* **213**, 461-462. Copyright 1981 de la AAAS.

39. Collins, R. L. (1970). The sound of one paw clapping: An inquiry into the origin of left-handedness. In Lindzey, G. y Thiessen, D. D. (eds.). *Contributions to Behavior-Genetic Analysis: The Mouse as Prototype*. Appleton-Century-Crofts.

40. Fawole, I. *op cit*. Estos datos son de la Tabla 3 del artículo.

41. Floersheim, G. L., Weber, O., Tschumi, P. y Ulbrich, M. (1983). Investigación citada en *Scientific American* **248** (abril 1983), n.º 4, p. 75.

42. Fuchs, J. A., Smith, J. D. y Bird, L. S. (1972). Genetic basis for an 11:5 dihybrid ratio observed in *Gossypium hirsutum*. *Journal of Heredity* **63**, 300-303. La base genética

de las razones 13:3 y 11:5 se explica en Strickberger, M. W. (1976). *Genetics*, 2.^a ed. New York: Macmillan, pp. 206-208.

43. Adaptado de Goodyear, C. P. (1970). Terrestrial and aquatic orientation in the starhead top-minnow, *Fundulus notii*. *Science* **168**, 603-605. Copyright 1970 de la AAAS.

44. Véase Batschelet, E. (1981). *Circular Statistics in Biology*. Academic Press.

45. Datos no publicados cortesía de J. L. Wolfson, recogidos en Bard College en 1997.

46. Paris, H. S. (1997). Genes for developmental fruit coloration of acorn squash. *Journal of Heredity* **88**, 52-56. El experimento incluía cruces de calabacines Table Queen (TQE) con Vegetable Spaghetti (VSP). Los datos presentados

en el ejercicio son de un cruce retrógrado de VSP con un cruce TQE $\times$ VSP.

47. Chen, P., Ma, G., Buss, G. R., Gunduz, I., Roane, C. W. y Tolin, S. A. (2001). Inheritance and alleism tests of Raiden soybean for resistance to soybean mosaic virus. *Journal of Heredity* **92**, 51-55. La clasificación «resistente» incluye plantas necróticas tanto resistentes como no resistentes.

48. Lamb, M. L., Fishbein, M., Douglas, J. M., Rhodes, F., Rogers, J., Bolan, G., Zenilman, J., Hoxworth, T., Malotte, K., Iatesta, M., Kent, C., Lentz, A., Graziano, S., Byers, R. H. y Peterman, T. A. (1998). Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases. *Journal of the American Medical Association* **280**, 1161-1167.

DATOS EN CATEGORÍAS: RELACIONES

10

Objetivos

En este capítulo ampliaremos nuestro estudio sobre datos en categorías a varias poblaciones:

- explicaremos la independencia y la asociación para variables categóricas;
- presentaremos un test chi-cuadrado para evaluar la independencia entre dos variables categóricas;
- consideraremos las condiciones de validez de un test chi-cuadrado;
- presentaremos el test exacto de independencia de Fisher entre dos variables categóricas;
- presentaremos el test de McNemar para analizar datos categóricos pareados;
- calcularemos el riesgo relativo, la razón de ventajas y su intervalo de confianza asociado.

10.1 Introducción

En el Capítulo 9 consideramos el análisis de una única muestra de datos categóricos. Las técnicas básicas que utilizamos fueron la estimación de las probabilidades de las categorías y la comparación de las frecuencias observadas de dichas categorías con las frecuencias «esperadas» de acuerdo con una hipótesis nula. En este capítulo ampliaremos estas técnicas básicas a situaciones más complicadas. Para plantear el escenario, presentaremos dos ejemplos. El primero de ellos presenta un experimento y el segundo un estudio observacional.

Ejemplo 10.1.1 Migrañas

Unos pacientes que padecían migrañas moderadas o severas participaron en un estudio clínico doble ciego para evaluar una cirugía experimental. Un grupo de 75 pacientes se asignó aleatoriamente para recibir cirugía real en los sitios de disparo de la migraña ($n = 49$) o una cirugía falsa de imitación ($n = 26$) en la que se realizaba una incisión pero ningún procedimiento posterior. Los cirujanos esperaban que los pacientes experimentaran «una reducción sustancial* de las migrañas», que etiquetaremos como «éxito». La Tabla 10.1.1 muestra los resultados del experimento¹.

Tabla 10.1.1 Respuesta a la cirugía para la migraña			
		Cirugía	
		Real	Falsa
¿Reducción sustancial de los dolores de cabeza?	Éxito	41	15
	Sin éxito	8	11
	Total	49	26

* «Reducción sustancial» significa al menos un 50 % de reducción de los dolores de cabeza en frecuencia, intensidad o duración cuando se compara con los valores del punto de partida (antes de la cirugía).

Una forma natural de expresar los resultados es utilizando porcentajes, como sigue:

- De las cirugías reales, $\frac{41}{49}$ o el 83,7 % tuvieron éxito.
- De las cirugías falsas de imitación, $\frac{15}{26}$ o el 57,7 % tuvieron éxito.

En este estudio la reducción sustancial en las migrañas fue más común en pacientes que recibieron cirugía real que en pacientes que recibieron la cirugía falsa de imitación (83,7 % frente a 57,7 %). La Tabla 10.1.2 presenta un resumen de los datos. La Figura 10.1.1 es un diagrama de barras que muestra los porcentajes de cirugías con éxito para los dos grupos. ■

Tabla 10.1.2 Respuesta a la cirugía para la migraña		
	Cirugía	
	Real	Falsa
<i>n</i>	49	26
Éxito	41	15
Porcentaje	83,7 %	57,7 %

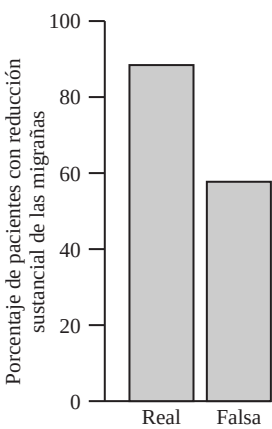


Figura 10.1.1 Diagrama de barras de los datos de cirugía para migrañas

Ejemplo 10.1.2 Test de VIH

En una muestra aleatoria de 120 estudiantes universitarios se encontró que 9 de 61 mujeres de la muestra habían realizado una prueba del VIH, en comparación con 8 de 59 hombres². Estos datos se muestran en la Tabla 10.1.3.

Tabla 10.1.3 Datos sobre la prueba del VIH		
	Mujeres	Hombres
Test del VIH	9	8
No test del VIH	52	51
Total	61	59

De las mujeres, $\frac{9}{61} = 0,148$ o el 14,8 % habían realizado la prueba. De los hombres, $\frac{8}{59} = 0,136$ o el 13,6 % habían realizado la prueba. Esos dos porcentajes son casi idénticos. ■

Las tablas como la 10.1.1 y la 10.1.3 se denominan **tablas de contingencia**. El foco de interés en una tabla de contingencia es la dependencia o asociación entre la variable de las columnas y la variable de las filas; por ejemplo, entre tratamientos y respuestas en las Tablas 10.1.1 y 10.1.3. (La palabra *contingente* significa «dependiente»). En particular, las Tablas 10.1.1 y 10.1.3 se denominan **tablas de contingencia de 2 × 2** («dos por dos»), ya que están formadas por dos

filas (excluyendo la fila de «total») y dos columnas. Cada categoría de la tabla de contingencia se denomina **celda**. Por tanto, una tabla de contingencia de 2×2 tiene cuatro celdas.

Consideraremos el análisis e interpretación de tablas de contingencia de 2×2 antes de ampliar el estudio a tablas mayores.

10.2 El test chi-cuadrado para la tabla de contingencia de 2×2

Cuando se analiza una tabla de contingencia de 2×2 es natural pensar en comparar la probabilidad de un suceso bajo una de las dos condiciones. Encontraremos útil ampliar el lenguaje de la probabilidad para que incluya un nuevo concepto: la probabilidad condicional*.

Probabilidad condicional

Recuérdese que la probabilidad de un suceso predice lo frecuentemente que ocurrirá dicho suceso. Una **probabilidad condicional** predice lo frecuentemente que ocurrirá un suceso bajo condiciones especificadas. La notación de una probabilidad condicional es

$$\Pr\{E|C\}$$

que se lee «probabilidad de E , dado C ». Cuando se estima una probabilidad condicional a partir de datos observados, la estimación se indica con el signo « $\hat{}$ », es decir,

$$\hat{\Pr}\{E|C\}$$

El siguiente ejemplo ilustra estas ideas.

Ejemplo 10.2.1 Migrañas

Consideremos los datos de migrañas del Ejemplo 10.1.1. Las probabilidades condicionales de interés son las siguientes:

$\Pr\{\text{reducción sustancial de migrañas}|\text{cirugía real}\} = \Pr\{\text{Éxito}|\text{Real}\} =$ probabilidad de que un paciente tenga una reducción sustancial de sus migrañas si ha recibido cirugía real.

$\Pr\{\text{reducción sustancial de migrañas}|\text{cirugía falsa de imitación}\} = \Pr\{\text{Éxito}|\text{Falsa}\} =$ probabilidad de que un paciente tenga una reducción sustancial de sus migrañas si ha recibido cirugía falsa de imitación.

Las estimaciones de estas probabilidades condicionales a partir de los datos de la Tabla 10.1.1 son

$$\hat{\Pr}(\text{Éxito}|\text{Real}) = \frac{41}{49} = 0,837$$

y

$$\hat{\Pr}(\text{Éxito}|\text{Falsa}) = \frac{15}{26} = 0,577$$

La hipótesis natural a contrastar es que las probabilidades condicionales asociadas con una tabla de 2×2 son iguales, que es lo mismo que decir que la probabilidad del suceso E no depende de si la primera condición, C , está presente o de si la segunda condición «no C », está presente.

$$H_0: \Pr\{E|C\} = \Pr\{E|\text{no } C\}$$

El siguiente ejemplo ilustra esta hipótesis nula.

* La probabilidad condicional se presenta también en la Sección opcional 3.3.

Ejemplo 10.2.2 Migrañas

En el estudio sobre migrañas del Ejemplo 10.1.1 la hipótesis nula es

$$H_0: \Pr\{\text{Éxito}|\text{Real}\} = \Pr\{\text{Éxito}|\text{Falsa}\}$$

o, de forma equivalente,

$$H_0: \Pr\{\text{Éxito}|\text{Real}\} = \Pr\{\text{Éxito}|\text{no Real}\}$$

El estadístico chi-cuadrado

Claramente, una forma natural de contrastar la anterior hipótesis nula será rechazar H_0 si $\hat{\Pr}\{E|C\}$ y $\hat{\Pr}\{E|\text{no } C\}$ son diferentes en una cantidad suficiente. Describimos un procedimiento de contraste que compara indirectamente con $\hat{\Pr}\{E|C\}$ y $\hat{\Pr}\{E|\text{no } C\}$, en vez de hacerlo directamente. El procedimiento es un test chi-cuadrado, basado en el estadístico χ_s^2 que se presentó en la Sección 9.4:

$$\chi_s^2 = \sum_{i=1}^4 \frac{(o_i - e_i)^2}{e_i}$$

En la fórmula, la suma se realiza sobre las cuatro celdas de la tabla de contingencia. Cada o representa la frecuencia observada y cada e representa la correspondiente frecuencia esperada de acuerdo con H_0 . Explicaremos seguidamente cómo calcular los valores de e .

El primer paso para determinar los e de una tabla de contingencia es calcular las frecuencias totales de las filas y las columnas (que se denominan también **frecuencias marginales**) y también la suma total de todas las frecuencias de las celdas. Los valores de e se deducen después de un razonamiento simple, como ilustra el Ejemplo 10.2.3.

Ejemplo 10.2.3 Migrañas

La Tabla 10.2.1 muestra los datos de migrañas del Ejemplo 10.1.1, junto con las frecuencias marginales.

Tabla 10.2.1 Frecuencias observadas en el estudio sobre migrañas			
	Cirugía		Total
	Real	Falsa	
Éxito	41	15	56
Sin éxito	8	11	19
Total	49	26	75

Los e deben ajustarse exactamente a la hipótesis nula. Como H_0 afirma que la probabilidad de éxito no depende del tratamiento, podemos generar una estimación de esta probabilidad uniendo los dos grupos de tratamiento. De la Tabla 10.2.1, la estimación conjunta, basada en el total marginal, es $\frac{56}{75}$. Es decir, si H_0 es cierta, entonces las colum-

nas «Real» y «Falsa» son equivalentes y podemos unir las. Nuestra mejor estimación de $\Pr\{\text{resultado de éxito}\}$ es $\frac{56}{75}$.

Podemos aplicar seguidamente este resultado a cada grupo de tratamiento para obtener el número esperado de resultados de éxito de acuerdo con H_0 , como sigue:

Grupo de cirugía real: $\frac{56}{75} \times 49 = 36,59$ resultados de éxito esperados.

Grupo de cirugía falsa: $\frac{56}{75} \times 26 = 19,41$ resultados de éxito esperados.

Asimismo, la estimación conjunta de $\Pr\{\text{la cirugía no tenga éxito}\}$ es $\frac{19}{75}$. Aplicando esta probabilidad a los dos grupos de tratamiento se obtiene

Grupo de cirugía real: $\frac{19}{75} \times 49 = 12,41$ resultados de fracaso esperados.

Grupo de cirugía falsa: $\frac{19}{75} \times 26 = 6,59$ resultados de fracaso esperados.

Las frecuencias esperadas se muestran entre paréntesis en la Tabla 10.2.2. Nótese que los totales marginales de las e son los mismos que los de las o . ■

Tabla 10.2.2 Frecuencias observadas y esperadas en el estudio sobre migrañas			
	Cirugía		Total
	Real	Falsa	
Éxito	41 (36,59)	15 (19,41)	56
Sin éxito	8 (12,41)	11 (6,59)	19
Total	49	26	75

En la práctica, no es necesario realizar esta cadena completa de razonamientos para obtener las frecuencias esperadas de una tabla de contingencia. El procedimiento para calcular las e se puede resumir en una simple fórmula. La frecuencia esperada de cada celda se calcula a partir de las frecuencias marginales totales de la fila y columna correspondientes, como sigue:

Frecuencias esperadas de una tabla de contingencia

$$e = \frac{(\text{Total de la fila}) \times (\text{Total de la columna})}{\text{Total completo}}$$

La fórmula produce el mismo resultado que el razonamiento dado en el Ejemplo 10.2.2, como muestra el ejemplo siguiente.

Ejemplo 10.2.4 Migrañas

Aplicaremos la fórmula anterior a los datos de migrañas del Ejemplo 10.1.1. La frecuencia esperada de resultados de éxito para la cirugía real se calcula a partir de los totales marginales como

$$e = \frac{56 \times 49}{75} = 36,59$$

Nótese que es el mismo resultado obtenido en el Ejemplo 10.2.2. Procediendo de forma similar para cada celda de la tabla de contingencia, se obtendrían los e que se muestran en la Tabla 10.2.2. ■

Nota: Aunque la fórmula de χ^2_s para tablas de contingencia es la misma que la del test de bondad de ajuste dada en la Sección 9.4, el método para calcular las e es bastante diferente en las tablas de contingencia porque la hipótesis nula es diferente.

El procedimiento de contraste

Aparte de las diferencias indicadas anteriormente al calcular los valores esperados, el test chi-cuadrado para una tabla de contingencia se calcula de forma similar al test chi-cuadrado de bondad de ajuste. Valores grandes de χ_s^2 indican evidencia en contra de H_0 . Los valores críticos se determinan mediante la Tabla 9. El número de grados de libertad de una tabla de contingencia de 2×2 es

$$gl = 1$$

El test chi-cuadrado de una tabla de 2×2 tiene un grado de libertad porque, en un cierto sentido, solo hay una celda libre en la tabla. La Tabla 10.2.2 tiene cuatro celdas, pero una vez hemos determinado que la frecuencia esperada de la celda superior izquierda es 36,59, la frecuencia esperada de la celda superior derecha está restringida a valer 19,41, ya que la suma total de las celdas de la fila superior debe ser 56. Asimismo, la celda inferior izquierda está restringida a valer 12,41, ya que la suma total de las celdas de la primera columna debe ser 49. Una vez esas tres celdas están determinadas, la celda inferior derecha queda también determinada. Por tanto, hay cuatro celdas en la tabla, pero solo una de ellas es «libre». Una vez hemos utilizado la hipótesis nula para determinar la frecuencia esperada de una celda, las otras celdas quedan determinadas.

En una tabla de contingencia de 2×2 , la hipótesis alternativa puede ser direccional o no direccional. Las alternativas direccionales se manejan mediante el procedimiento habitual en dos pasos, dividiendo por dos el P valor no direccional si los datos se desvían con respecto a H_0 en la dirección especificada por H_A (o indicando que el P valor es $> 0,50$ si los datos se desvían con respecto a H_0 en la dirección opuesta a la especificada por H_A). Nótese que χ_s^2 no expresa direccionalidad en sí mismo. Para calcular la direccionalidad de los datos, hay que calcular y comparar las probabilidades estimadas.

El ejemplo siguiente ilustra el test chi-cuadrado.

Ejemplo 10.2.5 Migrañas

En el experimento sobre migrañas del Ejemplo 10.1.1 apliquemos un test chi-cuadrado. Dado que el experimento requiere cirugía craneal, un error de Tipo I sería bastante serio y, por tanto, se requiere una elección conservativa de α . Utilizaremos $\alpha = 0,01$. Podemos plantear informalmente la hipótesis nula y una alternativa direccional de la forma siguiente:

H_0 : La cirugía real no es mejor que la falsa para reducir las migrañas.

H_A : La cirugía tradicional es mejor que la falsa para reducir las migrañas.

Utilizando la notación de probabilidades condicionales, las afirmaciones son:

$$H_0: \Pr\{\text{Éxito}|\text{Real}\} = \Pr\{\text{Éxito}|\text{Falsa}\}$$

$$H_A: \Pr\{\text{Éxito}|\text{Real}\} > \Pr\{\text{Éxito}|\text{Falsa}\}$$

Para comprobar la direccionalidad de los datos, calculamos las probabilidades de respuesta estimadas:

$$\hat{\Pr}\{\text{Éxito}|\text{Real}\} = \frac{41}{49} = 0,837$$

$$\hat{\Pr}\{\text{Éxito}|\text{Falsa}\} = \frac{15}{26} = 0,577$$

y podemos ver que

$$\hat{\Pr}\{\text{Éxito}|\text{Real}\} > \hat{\Pr}\{\text{Éxito}|\text{Falsa}\}$$

Por tanto, los datos se desvían con respecto a H_0 en la dirección especificada por H_A . Procedemos con el cálculo del estadístico chi-cuadrado para la Tabla 10.2.2:

$$\chi_s^2 = \frac{(41 - 36,59)^2}{36,59} + \frac{(15 - 19,41)^2}{19,41} + \frac{(8 - 12,41)^2}{12,41} + \frac{(11 - 6,59)^2}{6,59} = 6,06$$

Utilizando la Tabla 9 con $gl = 1$, encontramos que $\chi^2_{1,0,02} = 5,41$ y $\chi^2_{1,0,01} = 6,63$, y, por tanto, tenemos $0,005 < P \text{ valor} < 0,01$. Por tanto, rechazamos H_0 y encontramos que los datos proporcionan suficiente evidencia para concluir que la cirugía real es mejor que la cirugía falsa para reducir las migrañas.

Nótese que, incluso aunque $\hat{\Pr}\{\text{Reducción}|\text{Real}\}$ y $\hat{\Pr}\{\text{Reducción}|\text{Falsa}\}$ no intervengan en el cálculo de χ^2_s , el cálculo de $\hat{\Pr}\{\text{Reducción}|\text{Real}\}$ y $\hat{\Pr}\{\text{Reducción}|\text{Falsa}\}$ es una parte importante del procedimiento de contraste. La información proporcionada por los valores de $\hat{\Pr}\{\text{Reducción}|\text{Real}\}$ y $\hat{\Pr}\{\text{Reducción}|\text{Falsa}\}$ es esencial para una interpretación correcta de los resultados*.

Notas sobre el cálculo Los siguientes consejos son útiles para realizar una tabla de contingencia de 2×2 :

1. El formato de la tabla de contingencia es útil para los cálculos. Sin embargo, para presentar los datos en un informe, generalmente es mejor utilizar una forma más legible de presentación como la Tabla 10.1.2. En los ejercicios se muestran algunos ejemplos adicionales.
2. Para calcular χ^2_s , las frecuencias observadas (o) deben ser absolutas, y no relativas. Además, *la tabla debe contener las cuatro celdas*, de forma que la suma de los o sea igual al número total de observaciones.

Ilustración de la hipótesis nula

El estadístico chi-cuadrado mide la discrepancia entre los datos y la hipótesis nula de una forma indirecta. Las probabilidades condicionales muestrales intervienen de forma indirecta en el cálculo de las frecuencias esperadas. Si las probabilidades condicionales muestrales son iguales, entonces el valor de χ^2_s es cero. He aquí un ejemplo.

Ejemplo 10.2.6 Migrañas

La Tabla 10.2.3 muestra datos ficticios de un estudio sobre la migraña similar al descrito en el Ejemplo 10.1.1. Para los datos de la Tabla 10.2.3, las probabilidades estimadas de cirugía de éxito son iguales:

$$\hat{\Pr}\{\text{Éxito}|\text{Real}\} = \frac{30}{150} = 0,20$$

$$\hat{\Pr}\{\text{Éxito}|\text{Falsa}\} = \frac{20}{100} = 0,20$$

Podemos verificar fácilmente que, para la Tabla 10.2.3, las frecuencias esperadas son iguales a las frecuencias observadas, de forma que el valor de χ^2_s es cero. Nótese también que las columnas de la tabla son proporcionales entre sí:

$$\frac{30}{120} = \frac{20}{80}$$

Tabla 10.2.3 Datos ficticios de un estudio sobre la migraña			
	Cirugía		Total
	Real	Falsa	
Éxito	30	20	50
Sin éxito	120	80	200
Total	150	100	250

* Es natural preguntarse por qué no usar una comparación más directa de $\hat{\Pr}\{E|C\}$ y $\hat{\Pr}\{E|\text{no } C\}$. De hecho, hay un procedimiento de contraste basado en un estadístico de tipo t , que se calcula dividiendo $(\hat{\Pr}\{E|C\} - \hat{\Pr}\{E|\text{no } C\})$ por su error típico. Este procedimiento tipo t es equivalente al test chi-cuadrado. Hemos optado por presentar el test chi-cuadrado por dos razones: (1) se puede ampliar a tablas de contingencia mayores que 2×2 ; (2) en ciertas aplicaciones el test chi-cuadrado es más natural que el estadístico tipo t . Algunas de estas aplicaciones aparecen en la Sección 10.3.

Como sugiere el ejemplo anterior, un análisis «visual» de una tabla de contingencia se basa en comprobar la proporcionalidad de las columnas. Si las columnas son casi proporcionales, entonces los datos se ajustarán bastante bien a H_0 . Si son altamente no proporcionales, entonces los datos no estarán de acuerdo con H_0 . El siguiente ejemplo muestra un caso en el que los datos se ajustan bastante bien con las frecuencias esperadas bajo H_0 .

Ejemplo 10.2.7 Test de VIH

Los datos del Ejemplo 10.1.2 muestran porcentajes similares de hombres y mujeres que han realizado la prueba del VIH. La hipótesis nula natural es que $\Pr\{\text{Prueba del VIH}|\text{Mujer}\} = \Pr\{\text{Prueba del VIH}|\text{Hombre}\}$ y que las proporciones muestrales difieren solo debido al error por azar en el proceso de muestreo. Las frecuencias esperadas se muestran entre paréntesis en la Tabla 10.2.4. El estadístico de contraste chi-cuadrado es $\chi_s^2 = 0,035$. Utilizando la Tabla 8 con $gl = 1$, obtenemos que $\chi_{1,0,20}^2 = 1,64$. Por tanto, el P valor es mayor que 0,20 (utilizando un computador se obtiene P valor = 0,85) y no rechazamos la hipótesis nula. Nuestra conclusión es que los datos no proporcionan evidencia significativa a favor de que haya una diferencia en los porcentajes de hombres y mujeres que han realizado la prueba del VIH (en la universidad donde se realizó el estudio). ■

Tabla 10.2.4 Frecuencias observadas y esperadas en el estudio del VIH			
	Mujer	Hombre	Total
Prueba del VIH	9 (8,64)	8 (8,36)	17
No prueba del VIH	52 (52,36)	51 (50,64)	103
Total	61	59	120

Nótese que el valor real de χ_s^2 depende de los tamaños muestrales, así como del grado de no proporcionalidad. Como se comentó en la Sección 9.4, el valor de χ_s^2 varía directamente con el número de observaciones si la composición porcentual de los datos se mantiene fija y el número de observaciones varía. Esto refleja el hecho de que una desviación porcentual dada de H_0 es menos probable que ocurra por azar con un gran número de observaciones.

Ejercicios 10.2.1-10.2.14

10.2.1 La tabla de contingencia parcialmente completa que acompaña a este problema muestra las respuestas a dos tratamientos:

		Tratamiento	
		1	2
Respuesta	Éxito	70	
	Fracaso		
Total		100	200

- (a) Invente un conjunto de datos ficticio que esté de acuerdo con la tabla y para el que $\chi_s^2 = 0$.
 (b) Calcule las probabilidades estimadas de éxito ($\hat{\Pr}\{\text{Éxito}|\text{Tratamiento 1}\}$ y $\hat{\Pr}\{\text{Éxito}|\text{Tratamiento 2}\}$) para su conjunto de datos. ¿Son iguales?

10.2.2 Proceda como en el Ejercicio 10.2.1 para la siguiente tabla de contingencia:

		Tratamiento	
		1	2
Respuesta	Éxito	30	
	Fracaso		
Total		300	100

10.2.3 Proceda como en el Ejercicio 10.2.1 para la siguiente tabla de contingencia:

		Tratamiento	
		1	2
Respuesta	Éxito	5	20
	Fracaso	10	

10.2.4 La mayoría de las salamandras de la especie *P. cinerlus* tienen bandas rojas, pero algunos individuos son to-

talmente rojos. Se piensa que la forma totalmente roja es una imitación de la salamandra *N. viridescens*, que es tóxica para los pájaros. Para comprobar si la forma de imitación realmente sobrevive con más éxito, se expusieron a 163 ejemplares con bandas y 41 ejemplares rojos de *P. cinerlus* a la depredación de una población natural de pájaros. Después de dos horas, 65 individuos con bandas y 23 individuos rojos estaban todavía vivos³. Utilice un test chi-cuadrado para valorar la evidencia de que la forma de imitación sobrevive con más éxito. Utilice una alternativa direccional y $\alpha = 0,05$.

- Plantee la hipótesis nula con palabras.
- Plantee la hipótesis nula con símbolos.
- Calcule las proporciones de supervivencia de cada grupo y presente los valores en una tabla similar a la Tabla 10.1.2.
- Calcule el valor del estadístico de contraste y el P valor.
- Expresé la conclusión del test en el contexto de este ejercicio.

10.2.5 ¿Puede el ataque de un organismo a una planta inducir resistencia en la planta al ataque de un organismo diferente? En un estudio de esta cuestión, plantas de algodón (*Gossypium*) cultivadas en tiestos separados se asignaron aleatoriamente a dos grupos. Todas las plantas de un grupo se infectaron con un ácaro (*Tetranychus*). El otro grupo se mantuvo como control. Tras dos semanas se eliminaron los ácaros y se inoculó a las plantas *Verticillium*, un hongo causante de que la planta se marchite. La tabla que acompaña a este ejercicio muestra el número de plantas que desarrollaron síntomas de marchitamiento⁴. ¿Proporcionan los datos evidencia estadística suficiente para concluir que la infección con ácaros induce resistencia al marchitamiento por el hongo? Utilice un test chi-cuadrado con una alternativa direccional siguiendo los pasos (a)-(e) del Problema 10.2.4. Use $\alpha = 0,01$.

		Tratamiento	
		Ácaros	No ácaros
Respuesta	Marchitamiento	11	17
	No marchitamiento	15	4
	Total	26	21

10.2.6 Se ha sospechado que el uso continuado del teléfono móvil puede aumentar la probabilidad de desarrollar cáncer cerebral debido a la señal en frecuencia de microondas transmitida por el teléfono móvil. De acuerdo con esta teoría, si se mantiene de forma repetida un teléfono móvil cerca de un lado de la cabeza, es más probable desarrollar tumores cerebrales en ese lado. Para investigar esto, se estudió un grupo de pacientes que habían utilizado teléfonos móviles en los seis meses anteriores a desarrollar un tumor cerebral. Se preguntó a los pacientes si usaban de forma rutinaria el teléfono en uno de sus oídos, y si era así, cuál. Las 88 res-

puestas (de los que preferían un lado al otro) se muestran en la tabla siguiente⁵. ¿Proporcionan los datos evidencia estadística suficiente como para concluir que el uso de teléfonos móviles produce un incremento en los tumores cerebrales del lado correspondiente? Utilice un test chi-cuadrado contra una alternativa direccional siguiendo los cinco pasos (a)-(e) del Problema 10.2.4. Use $\alpha = 0,05$.

		Lado de uso del teléfono	
		Izquierdo	Derecho
Lado del tumor cerebral	Izquierdo	14	28
	Derecho	19	27
	Total	33	55

10.2.7 La fenitoína es un medicamento anticonvulsivo estándar que desafortunadamente presenta muchos efectos colaterales tóxicos. Se realizó un estudio para comparar la fenitoína con el valproato, otro medicamento utilizado en el tratamiento de la epilepsia. Se asignaron aleatoriamente una serie de pacientes para recibir fenitoína o valproato durante 12 meses. De los 20 pacientes que recibieron valproato, 6 estuvieron libres de ataques durante los 12 meses, mientras que 6 de los 17 pacientes que recibieron fenitoína estuvieron libres de ataques en ese mismo periodo⁶.

- Utilice un test chi-cuadrado para comparar las probabilidades de no sufrir un ataque para los dos medicamentos. Sea H_A la alternativa direccional y $\alpha = 0,10$. Siga los cinco pasos (a)-(e) del Problema 10.2.4.
- ¿Proporcionan sus conclusiones del apartado (a) evidencia de que el valproato y la fenitoína son igual de efectivos en la prevención de ataques? Comente su respuesta.

10.2.8 Se utilizan productos de sincronización del estro para producir el celo en las vacas de una forma predecible, de forma que puedan resultar embarazadas de forma fiable tras recibir inseminación artificial. En un estudio de dos productos de sincronización del estro, 42 vacas maduras (entre 4 y 8 años) se asignaron aleatoriamente para recibir el producto A o el producto B, y después, todas fueron inseminadas artificialmente. La tabla muestra cuántas inseminaciones resultaron en un embarazo⁷. Utilice un test chi-cuadrado para comparar la efectividad de los dos productos para producir un embarazo, utilizando los cinco pasos (a)-(e) del Problema 10.2.4. Utilice una alternativa no direccional y $\alpha = 0,05$.

	Tratamiento	
	Producto A	Producto B
Número total de vacas	21	21
Número de vacas embarazadas	8	15

10.2.9 En los estudios experimentales sobre el cáncer se usan a menudo variedades de animales que presentan una alta incidencia de tumores de forma natural. En un experimento de ese tipo, se mantuvieron en un entorno estéril ratones propensos a tumores. Un grupo de ratones se mantuvo completamente libre de gérmenes, y otro grupo se expuso a la bacteria intestinal *Escherichia coli*. La tabla siguiente muestra la incidencia de tumores de hígado⁸.

Tratamiento	Número total de ratones	Ratones con tumores de hígado	
		Número	Porcentaje
Libres de gérmenes	49	19	39 %
<i>E. coli</i>	13	8	62 %

- (a) ¿Qué fuerza tiene la evidencia de que la incidencia de tumores es mayor en ratones expuestos a *E. coli*? Utilice un test chi-cuadrado contra una alternativa direccional siguiendo los cinco pasos (a-e) del Problema 10.2.4. Utilice $\alpha = 0,05$.
- (b) ¿Cómo cambiaría el resultado del apartado (a) si los porcentajes (39 % y 62 %) de ratones con tumores fueran los mismos, pero los tamaños de la muestra se doblaran (98 y 26)? ¿Y si se triplicaran (149 y 39)? [Sugerencia: el apartado (a) casi no requiere cálculos].

10.2.10 En un estudio clínico aleatorizado para determinar los tiempos más efectivos de administración de medicamentos quimioterápicos a pacientes con cáncer de pulmón se administraron simultáneamente cuatro medicamentos a 16 pacientes, mientras que a otros 11 pacientes se les administraron de forma secuencial. Se observó la respuesta objetiva al tratamiento (definida como una reducción del tumor al menos en un 50 %) en 11 de los pacientes tratados simultáneamente y en 3 de los pacientes tratados secuencialmente⁹. ¿Proporcionan los datos evidencia sobre qué forma de administración es superior? Utilice un test chi-cuadrado contra una alternativa direccional siguiendo los cinco pasos (a-e) del Problema 10.2.4. Use $\alpha = 0,05$.

10.2.11 Unos médicos realizaron un experimento para investigar la efectividad de los protectores externos de cadera para evitar las fracturas de cadera en ancianos. Asignaron aleatoriamente a unas personas para que usaran protectores de cadera, mientras que otras se mantuvieron como grupo de control. Se apuntó el número de fracturas de cadera en cada grupo¹⁰. ¿Proporcionan los datos de la tabla siguiente evidencia suficiente para concluir que los protectores de cadera reducen la probabilidad de fractura? Utilice un test chi-cuadrado contra una alternativa direccional siguiendo los cinco pasos (a-e) del Problema 10.2.4. Utilice $\alpha = 0,01$.

		Tratamiento	
		Protector de cadera	Control
Respuesta	Fractura de cadera	13	67
	No fractura de cadera	640	1.081
	Total	653	1.148

10.2.12 Una muestra de 276 voluntarios adultos sanos fueron entrevistados sobre la variedad de relaciones sociales que tenían (por ejemplo, relaciones con los padres, vecinos cercanos, compañeros de trabajo, etc.). Se les administraron gotas nasales con un rinovirus y se les mantuvo en cuarentena cinco días. De los 123 sujetos que estaban en cinco tipos o menos de relaciones sociales, 57 (46,3 %) contrajeron un catarro. De los 153 que estaban en al menos seis tipos de relaciones sociales, 52 (34,0 %) contrajeron un catarro¹¹. Por tanto, los datos sugieren que tener más tipos de relaciones sociales ayuda a desarrollar resistencia al catarro común. Determine si esta diferencia es estadísticamente significativa. Es decir, utilice un test chi-cuadrado para contrastar la hipótesis nula de que la probabilidad de contraer un catarro no depende del número de relaciones sociales de una persona, siguiendo los cinco pasos (a-e) del Problema 10.2.4. Utilice una alternativa no direccional y $\alpha = 0,05$.

10.2.13 El medicamento ancrod se experimentó en un ensayo clínico doble ciego en el que sujetos con derrame cerebral se asignaron aleatoriamente para tomar ancrod o un placebo. Una de las variables de respuesta del estudio fue si el sujeto sufrió o no hemorragia intracraneal¹². Los datos se muestran en la tabla siguiente. Utilice un test chi-cuadrado para determinar si la diferencia en los porcentajes de hemorragias es estadísticamente significativa, siguiendo el procedimiento de cinco pasos (a-e) del Problema 10.2.4. Utilice una alternativa no direccional y $\alpha = 0,05$.

		Tratamiento	
		Ancrod	Placebo
¿Hemorragia?	Sí	13	5
	No	235	247
	Total	248	252

10.2.14 ¿Responden mejor las mujeres a las peticiones de los hombres durante la fase fértil de sus ciclos menstruales? En un estudio sobre esta cuestión 200 mujeres entre 18 y 25 años que caminaban solas por una ciudad fueron abordadas por un hombre atractivo de 20 años que les pidió su número de teléfono. Estudios previos sugerían que durante la fase fértil de su ciclo menstrual una mujer podría ser más recepti-

va a esta clase de petición que en otros momentos. De las 60 mujeres que estaban en la fase fértil de sus ciclos 13 dieron su número de teléfono y 47 rehusaron. Los números correspondientes para las 140 mujeres que no estaban en su fase fértil fueron de 11 y 129¹³. Los datos se resumen en la tabla siguiente. Utilice un test chi-cuadrado para determinar si la diferencia en las probabilidades de éxito proporciona evidencia significativa a favor de una alternativa direccional apropiada, empleando $\alpha = 0,02$ y siguiendo el procedimiento de cinco pasos (a-e) del Problema 10.2.4.

		Fase	
		Fértil	No
¿Éxito?	Sí	13	11
	No	47	129
	Total	60	140

10.3 Independencia y asociación en la tabla de contingencia de 2×2

La tabla de contingencia de 2×2 es aparentemente simple. En esta sección exploraremos relaciones adicionales que puede expresar.

Dos contextos para las tablas de contingencia

Una tabla de contingencia de 2×2 puede surgir en dos contextos, concretamente:

1. Dos muestras independientes con una variable observada dicotómica.
2. Una muestra con dos variables dicotómicas observadas.

El primer contexto es el que ilustran los datos sobre migrañas del Ejemplo 10.1.1, que se pueden ver como dos muestras independientes (el grupo de cirugía real y el grupo de cirugía falsa) de tamaños $n_1 = 49$ y $n_2 = 26$. La variable observada es el éxito (o fracaso) de la cirugía. Cualquier estudio donde intervenga una variable dicotómica y se realice una asignación completamente aleatoria de dos tratamientos se puede ver de esta forma. El segundo contexto es el que ilustran los datos sobre el VIH del ejemplo 10.1.2, que se pueden ver como una única muestra de $n = 120$ estudiantes observada con respecto a dos variables dicotómicas: Sexo (hombre o mujer) y estado con respecto a la prueba del VIH (si el estudiante ha realizado o no la prueba del VIH).

Los dos contextos (dos muestras con una variable o una muestra con dos variables) muchas veces no se diferencian claramente. Por ejemplo, los datos sobre el VIH del Ejemplo 10.1.2 se podrían haber tomado en dos muestras (61 mujeres y 59 hombres) observadas con respecto a una variable dicotómica (estado con respecto a la prueba del VIH).

La aritmética del test chi-cuadrado es la misma en ambos contextos, pero el planteamiento y la interpretación de las hipótesis y conclusiones puede ser muy diferente.

Independencia y asociación

En muchas tablas de contingencia, las columnas de la tabla tienen un papel diferente al de las filas. Por ejemplo, en los datos sobre migrañas del Ejemplo 10.1.1, las columnas representan tratamientos y las filas representan respuestas. Asimismo, en el Ejemplo 10.1.2 parece más natural definir las probabilidades condicionales por columnas $\Pr\{\text{Prueba VIH}|\text{M}\}$ y $\Pr\{\text{Prueba VIH}|\text{H}\}$ en vez de las probabilidades condicionales por filas $\Pr\{\text{M}|\text{Prueba VIH}\}$ y $\Pr\{\text{H}|\text{Prueba VIH}\}$.

Por otra parte, en algunos casos es natural ver con papeles intercambiables a las filas y las columnas de la tabla de contingencia. La hipótesis nula del test chi-cuadrado se puede expresar entonces por filas o por columnas. Sigue a continuación un ejemplo.

Ejemplo 10.3.1 Color de pelo y color de ojos

Para estudiar la relación entre el color de pelo y el color de ojos en la población alemana, un antropólogo observó una muestra de 6.800 hombres, con los resultados que se muestran en la Tabla 10.3.1¹⁴.

Los datos de la Tabla 10.3.1 se podrían ver de forma natural como una única muestra de tamaño $n = 6.800$ con dos variables dicotómicas observadas: color de pelo y color de ojos. Para describir los datos, denotaremos ojos

Tabla 10.3.1 Color de pelo y color de ojos				
		Color de pelo		Total
		Oscuro	Claro	
Color de ojos	Oscuro	726	131	857
	Claro	3.129	2.814	5.943
	Total	3.855	2.945	6.800

oscuros o claros como OO y OC, y pelo oscuro o claro como PO y PC. Podemos calcular las probabilidades condicionales estimadas por columnas como sigue:

$$\hat{\Pr}\{OO|PO\} = \frac{726}{3.855} \approx 0,19$$

$$\hat{\Pr}\{OO|PC\} = \frac{131}{2.945} \approx 0,04$$

Una forma natural de analizar los datos es comparar estos valores: 0,19 frente a 0,04. Por otra parte, es igual de natural calcular y comparar las probabilidades condicionales estimadas por filas:

$$\hat{\Pr}\{PO|OO\} = \frac{726}{857} \approx 0,85$$

$$\hat{\Pr}\{PO|OC\} = \frac{3.129}{5.943} \approx 0,53$$

Correspondiéndose con las dos perspectivas de la tabla de contingencia, la hipótesis nula del test chi-cuadrado por columnas se podría plantear como

$$H_0: \Pr\{OO|PO\} = \Pr\{OO|PC\}$$

o por filas como

$$H_0: \Pr\{PO|OO\} = \Pr\{PO|OC\}$$

Como veremos, las dos hipótesis son equivalentes. Es decir, cualquier población que cumpla una cumplirá también la otra. ■

Cuando un conjunto de datos se ve como una única muestra con dos variables observadas, la relación expresada por H_0 se denomina **independencia estadística** de la variable de las filas y la variable de las columnas. Las variables que no son independientes se denominan **dependientes** o **asociadas**. Por tanto, el test chi-cuadrado se denomina a veces «test de independencia» o «test de asociación».

Ejemplo 10.3.2 Color de pelo y color de ojos

La hipótesis nula del Ejemplo 10.3.1 se puede expresar verbalmente como

H_0 : El color de ojos es independiente del color de pelo

o

H_0 : El color de pelo es independiente del color de ojos

o, de forma más simétrica

H_0 : El color de pelo y el color de ojos son independientes ■

La hipótesis nula de independencia se puede plantear de forma genérica como sigue. Se van a comparar dos grupos G_1 y G_2 con respecto a la probabilidad de una característica C . La hipótesis nula es

$$H_0: \Pr\{C|G_1\} = \Pr\{C|G_2\}$$

Nótese que todas las afirmaciones de H_0 en el Ejemplo 10.3.1 son de esta forma.

Para clarificar más el significado de la hipótesis nula de independencia, en el ejemplo siguiente examinamos un conjunto de datos que se ajusta *exactamente* a H_0 .

Ejemplo 10.3.3 Altura de plantas y resistencia a enfermedades

Consideremos una especie (ficticia) de planta que se puede clasificar en baja (B) y alta (A) y resistente (R) o no resistente (NR) a una cierta enfermedad. Consideremos la siguiente hipótesis nula:

H_0 : La altura de la planta y la resistencia a la enfermedad son independientes

Todos los planteamientos siguientes de H_0 son válidos:

1. $H_0: \Pr\{R|B\} = \Pr\{R|A\}$
2. $H_0: \Pr\{NR|B\} = \Pr\{NR|A\}$
3. $H_0: \Pr\{B|R\} = \Pr\{B|NR\}$
4. $H_0: \Pr\{A|R\} = \Pr\{A|NR\}$

El siguiente no es un planteamiento de H_0 :

5. $H_0: \Pr\{R|B\} = \Pr\{NR|B\}$

Nótese la diferencia entre los planteamientos 5 y 1. El planteamiento 1 compara dos grupos (plantas bajas y altas) con respecto a su resistencia a la enfermedad, mientras que el planteamiento 5 es un planteamiento sobre la distribución de la resistencia a la enfermedad en solo *un* grupo (plantas bajas). El planteamiento 5 dice simplemente que el 50 % de las plantas bajas son resistentes y el 50 % no resistentes.

Supongamos, ahora, que tomamos una muestra aleatoria de 100 plantas de la población y obtenemos los datos de la Tabla 10.3.2.

Tabla 10.3.2 Altura de plantas y resistencia a una enfermedad				
		Altura		Total
		B	A	
Resistencia	R	12	18	30
	NR	28	42	70
	Total	40	60	100

Los datos de la Table 10.3.2 se ajustan exactamente a H_0 . Este ajuste se puede comprobar de cuatro formas distintas, correspondientes a los cuatro planteamientos simbólicos de H_0 :

1. $\hat{\Pr}\{R|B\} = \hat{\Pr}\{R|A\}$

$$\frac{12}{40} = 0,30 = \frac{18}{60}$$

2. $\hat{\Pr}\{NR|B\} = \hat{\Pr}\{NR|A\}$

$$\frac{28}{40} = 0,70 = \frac{42}{60}$$

$$3. \hat{\Pr}\{B|R\} = \hat{\Pr}\{B|NR\}$$

$$\frac{12}{30} = 0,40 = \frac{28}{70}$$

$$4. \hat{\Pr}\{A|R\} = \hat{\Pr}\{A|NR\}$$

$$\frac{18}{30} = 0,60 = \frac{42}{70}$$

Nótese que los datos de la Tabla 10.3.2 *no* se ajusta al planteamiento 5:

$$\hat{\Pr}\{R|B\} = \frac{12}{40} = 0,30 \text{ y } \hat{\Pr}\{NR|B\} = \frac{28}{40} = 0,70$$

$$0,30 \neq 0,70$$



Propiedades sobre filas y columnas

Los datos de la Tabla 10.3.2 muestran independencia tanto si se observan por filas como por columnas. Esto no es un casual, como muestra el hecho siguiente.

Propiedad 10.3.1 Las columnas de una tabla 2×2 son proporcionales si y solo si las filas son proporcionales. Concretamente, supongamos que a , b , c y d son números positivos cualesquiera, dispuestos como indica la Tabla 10.3.3.

Tabla 10.3.3 Una tabla de contingencia 2×2 genérica			
			Total
	a	b	$a + b$
	c	d	$c + d$
Total	$a + c$	$b + d$	

Entonces

$$\frac{a}{c} = \frac{b}{d} \text{ si y solo si } \frac{a}{b} = \frac{c}{d}$$

Otra forma de expresarlo es

$$\frac{a}{a+c} = \frac{b}{b+d} \text{ si y solo si } \frac{a}{a+b} = \frac{c}{c+d}$$

Se puede demostrar fácilmente que la Propiedad 10.3.1 es cierta. Basta con utilizar álgebra elemental. Debido a la Propiedad 10.3.1, la relación de independencia en una tabla de contingencia de 2×2 es la misma tanto si la tabla se ve por filas como si se ve por columnas. Nótese también que las frecuencias esperadas y, por tanto, el valor de χ^2_s , permanecería igual si se intercambiaran las filas y las columnas de la tabla de contingencia.

La propiedad siguiente muestra que la *dirección* de dependencia es la misma tanto por filas como por columnas.

Propiedad 10.3.2 Supongamos que a , b , c y d son números positivos cualesquiera, dispuestos como indica la Tabla 10.3.3. Entonces

$$\frac{a}{a+c} > \frac{b}{b+d} \text{ si y solo si } \frac{a}{a+b} > \frac{c}{c+d}$$

Igualmente,

$$\frac{a}{a+c} < \frac{b}{b+d} \quad \text{si y solo si} \quad \frac{a}{a+b} < \frac{c}{c+d}$$

Nota: Para una explicación adicional sobre probabilidad condicional e independencia, véase la Sección opcional 3.3.

Descripción verbal de la asociación

Las ideas de implicación lógica se expresan todos los días en castellano de formas sutiles. El siguiente extracto es de *Alicia en el País de las Maravillas*, de Lewis Carroll.

«... deberías decir lo que quieres decir», continuó la Liebre de Marzo.

«Lo hago», replicó Alicia apresuradamente; «al menos —al menos quiero decir lo que digo— es lo mismo, como sabes».

«¡Casi que no es lo mismo!» dijo el Sombrero Loco. «Porque entonces, ¿se podría decir que “como lo que veo” es lo mismo que “veo lo que como”!»

... «También podrías decir», añadió el Lirón, «¡Que “respiro cuando duermo” es lo mismo que “duermo cuando respiro”!»

«Es lo mismo para ti», dijo el Sombrero Loco.

Nosotros usamos también el lenguaje ordinario para expresar ideas de probabilidad, de probabilidad condicional y de asociación. Por ejemplo, consideremos las cuatro afirmaciones siguientes:

La ceguera al color es más común en hombres que en mujeres.

La masculinidad es más común que la feminidad en personas ciegas al color.

La mayoría de las personas ciegas al color son hombres.

La mayoría de los hombres son ciegos al color.

Las tres afirmaciones primeras son verdaderas. Realmente son formas diferentes de decir lo mismo. Sin embargo, la última afirmación es falsa¹⁵.

Al interpretar las tablas de contingencia, a menudo es necesario expresar con palabras las relaciones probabilísticas. Puede ser todo un desafío. Si conseguimos tener fluidez en esa descripción, entonces siempre «diremos lo que queremos decir» y «querremos decir lo que decimos». Los dos ejemplos siguientes ilustran algunos de estos aspectos.

Ejemplo 10.3.4 Altura de plantas y resistencia a enfermedades

En el estudio sobre altura y resistencia a enfermedades de plantas del Ejemplo 10.3.3 consideramos la hipótesis nula

H_0 : la altura y la resistencia son independientes.

Esta hipótesis se podría expresar también verbalmente de algunas otras formas, como

H_0 : las plantas bajas y las altas tienen igual probabilidad de ser resistentes.

H_0 : las plantas bajas y altas tienen la misma probabilidad de ser resistentes.

H_0 : la resistencia es igualmente común en plantas bajas y altas. ■

Ejemplo 10.3.5 Color de pelo y color de ojos

Consideremos la interpretación de la Tabla 10.3.1. El estadístico chi-cuadrado es $\chi_s^2 = 314$. Utilizando la Tabla 9 vemos que el P valor es muy pequeño, por lo que la hipótesis de independencia debe ser contundentemente rechazada. Podríamos plantear nuestra conclusión de varias formas. Por ejemplo, supongamos que nos enfocamos en la incidencia de ojos oscuros. A partir de los datos encontramos que

$$\hat{\Pr}\{\text{OO}|\text{PO}\} > \hat{\Pr}\{\text{OO}|\text{PC}\}$$

es decir,

$$\frac{726}{3.855} = 0,19 > \frac{131}{2.945} = 0,04$$

Una conclusión natural de esta comparación sería:

Conclusión 1: Hay evidencia suficiente para concluir que los hombres de pelo oscuro tienen una mayor tendencia que los hombres de pelo claro a tener ojos oscuros.

Esta afirmación se ha construido cuidadosamente, porque la afirmación

«Los hombres con pelo oscuro tienen una tendencia mayor a tener ojos oscuros».

es ambigua en sí misma. Podría significar

«Los hombres de pelo oscuro tienen una tendencia mayor que los hombres de pelo claro a tener ojos oscuros».

o

«Los hombres con pelo oscuro tienen una tendencia mayor a tener ojos oscuros que a tener ojos claros».

La primera de estas afirmaciones dice que

$$\hat{\Pr}\{OO|PO\} > \hat{\Pr}\{OO|PC\}$$

Mientras que la segunda dice que

$$\hat{\Pr}\{OO|PO\} > \hat{\Pr}\{OC|PO\}$$

La segunda afirmación asevera que más de la mitad de los hombres con pelo oscuro tienen ojos oscuros. Nótese que los datos no fundamentan esta aseveración. De los 3.855 hombres con pelo oscuro solo el 19 % tienen ojos oscuros.

La conclusión 1 solo es una de las formas posibles de expresar la conclusión que se obtiene del análisis de la tabla de contingencia. Por ejemplo, alguien se podría enfocar en el pelo oscuro y expresar:

Conclusión 2: Existe evidencia suficiente para concluir que los hombres de ojos oscuros tienen una tendencia mayor a tener cabello oscuro que los hombres de ojos claros.

Una forma más simétrica de expresarlo sería

Conclusión 3: Existe evidencia suficiente para concluir que el cabello oscuro está asociado con ojos oscuros.

Sin embargo, la forma de expresar la conclusión 3 se puede malinterpretar fácilmente. Podría sugerir algo como:

«Existe evidencia suficiente para concluir que la mayoría de los hombres con pelo oscuro tienen ojos oscuros»
que no es una interpretación correcta. ■

Debemos resaltar de nuevo el principio que establecimos en la Sección 10.2: *el cálculo y la comparación de las probabilidades condicionales o apropiadas es una parte esencial del test chi-cuadrado*. El Ejemplo 10.3.5 proporciona una amplia ilustración de este punto.

Ejercicios 10.3.1-10.3.12

10.3.1 Considere una población ficticia de ratones. El pelaje de cada animal es de color negro (N) o gris (G) y de textura ondulada (O) o suave (S). Expresar cada una de las siguientes relaciones en términos de probabilidad o probabilidades condicionales relacionadas con la población de animales.

(a) Los pelajes suaves son más comunes entre los ratones negros que entre los ratones grises.

(b) Los pelajes suaves son más comunes que los pelajes ondulados entre los ratones negros.

(c) Los pelajes suaves son más a menudo negros que los pelajes ondulados.

(d) Los pelajes suaves son más a menudo negros que grises.

(e) Los pelajes suaves son más comunes que los pelajes ondulados.

10.3.2 Considere una población ficticia de ratones en la que el pelaje de cada animal es de color negro (N) o gris (G) y de textura ondulada (O) o suave (S) (como en el Ejercicio 10.3.1). Suponga que se selecciona una muestra aleatoria de ratones de la población y que se observa el color y la textura del pelaje. Considere la tabla de contingencia de los datos parcialmente completa que acompaña a este ejercicio.

		Color	
		N	G
Textura	O		50
	S		
	Total	60	150

(a) Invente conjuntos de datos ficticios que se ajusten a la tabla y para los que

(i) $\hat{\Pr}\{O|N\} > \hat{\Pr}\{O|G\}$; (ii) $\hat{\Pr}\{O|N\} = \hat{\Pr}\{O|G\}$

En cada caso, verifique su respuesta calculando las probabilidades condicionales estimadas.

(b) Para cada uno de los dos conjuntos de datos que inventó en el apartado (a), calcule $\hat{\Pr}\{N|O\}$ y $\hat{\Pr}\{N|S\}$.

(c) ¿Cuál de los conjuntos de datos del apartado (a) tiene $\hat{\Pr}\{N|O\} > \hat{\Pr}\{N|S\}$? Indique si se podría inventar un conjunto de datos para el que

$\hat{\Pr}\{O|N\} > \hat{\Pr}\{O|G\}$ pero $\hat{\Pr}\{O|N\} < \hat{\Pr}\{O|S\}$

Si es posible, hágalo. Si no es posible, explique por qué.

10.3.3 Una serie de hombres con cáncer de próstata se asignaron aleatoriamente a recibir cirugía ($n = 347$) o a «espera vigilante» (no cirugía, $n = 348$). En los siguientes años hubo 83 muertes en el primer grupo y 106 muertes en el segundo grupo. Los resultados se muestran en la tabla¹⁶.

		Tratamiento		Total
		Cirugía	EV	
Supervivencia	Fallecidos	83	106	189
	Vivos	264	242	506
	Total	347	348	695

(a) Sean F y V las letras que representan a los fallecidos y a los vivos, respectivamente, y sean C y EV las letras que representan a la cirugía y a la espera vigilante. Estime $\Pr\{F|C\}$ y $\Pr\{F|EV\}$.

(b) El valor del estadístico chi-cuadrado de la tabla de contingencia para estos datos es $\chi_s^2 = 3,75$. Contraste la relación entre el tratamiento y la supervivencia. Utilice una alternativa no direccional y $\alpha = 0,05$.

10.3.4 En un estudio sobre las simetrías del comportamiento se preguntó a 2.391 mujeres qué mano preferían usar (por ejemplo, para escribir) y qué pie preferían usar (por ejemplo, para golpear una pelota). Los resultados se muestran en la tabla¹⁷.

Mano preferida	Pie preferido	Número de mujeres
Derecha	Derecho	2.012
Derecha	Izquierdo	142
Izquierda	Derecho	121
Izquierda	Izquierdo	116
	Total	2.391

(a) Estime la probabilidad condicional de que una mujer utilice el pie derecho, dado que utiliza la mano derecha.

(b) Estime la probabilidad condicional de que una mujer utilice el pie derecho, dado que utiliza la mano izquierda.

(c) Suponga que se desea contrastar la hipótesis nula de que la preferencia de la mano y la preferencia del pie son independientes. Calcule el estadístico chi-cuadrado para esta hipótesis.

(d) Suponga que se desea contrastar la hipótesis nula de que es igual de probable que las mujeres que utilizan la mano derecha utilicen el pie derecho o el pie izquierdo. Calcule el estadístico chi-cuadrado para esta hipótesis.

10.3.5 Considere un estudio para investigar un cierto agente sospechoso de causar una enfermedad. Se escogen aleatoriamente 1.000 personas de una población, y cada individuo se clasifica como enfermo o no enfermo y como expuesto o no expuesto a la gente. Los resultados se van a representar en la siguiente tabla de contingencia:

		Exposición	
		Sí	No
Enfermedad	Sí		
	No		

Supongamos que ES y EN indican exposición y no exposición y FS y FN indican la presencia y ausencia de la enfermedad. Exprese cada una de las siguientes afirmaciones en términos de probabilidades condicionales. (Considere que «una mayoría» significa «más de la mitad»).

(a) La enfermedad es más común entre las personas expuestas que entre las personas no expuestas.

(b) La exposición es más común entre las personas enfermas que entre las personas no enfermas.

(c) La exposición es más común entre las personas enfermas que la no exposición.

(d) Una mayoría de las personas enfermas son expuestas.

(e) Una mayoría de las personas expuestas están enfermas.

(f) Es más probable que estén enfermas las personas expuestas que las no expuestas.

(g) Es más probable que las personas expuestas estén enfermas que el que no estén enfermas.

10.3.6 En referencia al Ejercicio 10.3.5, ¿cuál de las afirmaciones expresa la aserción de que la ocurrencia de la enfermedad está asociada con la exposición al agente? (Puede haber más de una).

10.3.7 En referencia al Ejercicio 10.3.5, invente los conjuntos de datos ficticios que se especifican posteriormente, y verifique su respuesta calculando las probabilidades condicionadas estimadas apropiadas. (No es necesario que sus datos sean estadísticamente significativos).

(a) Invente un conjunto de datos para los que

$$\hat{Pr}\{FS|ES\} > \hat{Pr}\{FS|EN\} \text{ pero}$$

$$\hat{Pr}\{ES|FS\} < \hat{Pr}\{EN|FS\}$$

o explique por qué no es posible.

(b) Invente un conjunto de datos que esté de acuerdo con la afirmación (a) del Ejercicio 10.3.5, pero que no esté de acuerdo con (d) ni con (e). Si no es posible, explique por qué.

(c) Invente un conjunto de datos para los que

$$\hat{Pr}\{FS|ES\} > \hat{Pr}\{FS|EN\} \text{ pero}$$

$$\hat{Pr}\{ES|FS\} < \hat{Pr}\{EN|FN\}$$

o explique por qué no es posible.

10.3.8 Un ecólogo estudió la distribución espacial de especies de árboles en un área de bosque. De un área total de 25 acres, seleccionó aleatoriamente 144 cuadrantes (parcelas) cada una de ellas de 38 pies cuadrados, y anotó la presencia o ausencia de arces y nogales americanos en cada cuadrante. Los resultados se muestran en la tabla¹⁸.

		Arces	
		Presentes	Ausentes
Nogales americanos	Presentes	26	63
	Ausentes	29	26

El valor del estadístico chi-cuadrado para esta tabla de contingencia es $\chi_s^2 = 7.96$. Contraste la hipótesis nula de que las dos especies están distribuidas independientemente entre sí. Utilice una alternativa no direccional y $\alpha = 0.01$. Al plantear su conclusión, indique si los datos sugieren atracción o repulsión entre las especies. Fundamente su interpretación con las probabilidades condicionadas estimadas a partir de los datos.

10.3.9 En referencia al Ejercicio 10.3.8, suponga que los datos de dos especies ficticias de árboles, A y B, son los que se presentan en la tabla que acompaña a este ejercicio. El valor del estadístico chi-cuadrado para esta tabla de contingencia es $\chi_s^2 = 9.07$. Como en el Ejercicio 10.3.8, contraste la hipótesis nula de independencia e interprete la conclusión en términos de atracción o repulsión entre las especies.

		Especie A	
		Presentes	Ausentes
Especie B	Presentes	30	10
	Ausentes	49	55

10.3.10 Se realizó un experimento aleatorio en el que una serie de pacientes con enfermedad coronaria recibieron angioplastia o cirugía de *bypass*. La tabla que acompaña a este ejercicio muestra la incidencia de angina de pecho entre los pacientes cinco años después del tratamiento¹⁹.

		Tratamiento		Total
		Angioplastia	Bypass	
¿Angina?	Sí	111	74	185
	No	402	441	843
	Total	513	515	1.028

Sea A la letra que representa la angioplastia y B la letra que representa el *bypass*.

(a) Calcule $\hat{Pr}\{S|A\}$ y $\hat{Pr}\{S|B\}$.

(b) Calcule $\hat{Pr}\{A|S\}$ y $\hat{Pr}\{A|No\}$.

10.3.11 En referencia al Ejercicio 10.3.10, invente un conjunto de datos ficticio sobre el tratamiento coronario y la angina de pecho para 1.000 pacientes, para los que $\hat{Pr}\{S|A\}$ sea dos veces mayor que $\hat{Pr}\{S|B\}$, pero que no obstante la mayoría de los pacientes que tienen angina de pecho han recibido cirugía de *bypass* (y no angioplastia).

10.3.12 Suponga que se examinan parejas de hermanos gemelos para ver la mano que utiliza cada uno de ellos. Suponga que todos los gemelos son parejas hermano-hermana. Suponga que se obtienen datos de 1.000 parejas de gemelos, con los resultados que se muestran en la tabla siguiente²⁰. Indique si cada una de las siguientes afirmaciones es verdadera o falsa.

- La mayoría de los hermanos usan la misma mano que sus hermanas.
- La mayoría de las hermanas usan la misma mano que sus hermanos.
- En la mayoría de las parejas de gemelos ambos usan o bien la mano derecha o la mano izquierda.
- La mano que usa una hermana gemela es independiente de la mano que usa su hermano gemelo.
- La mayoría de las hermanas que usan la mano izquierda tienen hermanos que usan la mano derecha.

		Hermana		Total
		Izquierda	Derecha	
Hermano	Izquierda	15	85	100
	Derecha	135	765	900
	Total	150	850	1.000

10.4 Test exacto de Fisher (opcional)

En esta sección opcional consideraremos una alternativa al test chi-cuadrado para tablas de contingencia 2×2 . Este procedimiento, conocido como **test exacto de Fisher** es particularmente apropiado cuando se está tratando con muestras pequeñas. El Ejemplo 10.4.1 presenta una situación en la cual se puede utilizar el test exacto de Fisher.

Ejemplo 10.4.1 OMEC

La oxigenación mediante membrana extracorpórea (OMEC) es un procedimiento que potencialmente puede salvar la vida y que se utiliza para tratar a bebés recién nacidos que padecen fallo respiratorio severo. Se realizó un experimento con 29 bebés que fueron tratados con OMEC y 10 bebés que fueron tratados con la terapia médica convencional (TMC). Los datos se muestran en la Tabla 10.4.1²¹.

Tabla 10.4.1 Datos del experimento de OMEC				
		Tratamiento		Total
		TMC	OMEC	
Resultado	Fallecen	4	1	5
	Sobreviven	6	28	34
	Total	10	29	39

Los datos de la Tabla 10.4.1 muestran que 34 de los 39 bebés sobrevivieron, pero 5 de ellos fallecieron. El porcentaje de fallecimientos fue del 40 % en los que recibieron TMC y del 3,4 % para los que recibieron OMEC. Sin embargo, los tamaños muestrales son bastante pequeños. ¿Es posible que la diferencia en los porcentajes de fallecimientos fuera debida únicamente al azar?

La hipótesis nula de interés es que el resultado (sobrevivir o fallecer) es independiente del tratamiento (TMC o OMEC). Si la hipótesis nula es cierta, entonces podemos ver los datos de la siguiente forma: los dos encabezamientos de las columnas «TMC» y «OMEC» son etiquetas arbitrarias. 5 bebés podían haber fallecido independientemente de en qué grupo de tratamiento estaban; 4 de esos bebés terminaron en el grupo TMC por azar.

La hipótesis alternativa establece que la probabilidad de fallecimiento depende del grupo de tratamiento. Esto significa que existe una diferencia real entre los porcentajes de supervivencia de TMC y OMEC, y eso es lo que hace que los porcentajes muestrales sean diferentes.

Por tanto, una pregunta de interés es esta: «Si la hipótesis nula es cierta, ¿como es de probable obtener una tabla de datos como la Tabla 10.4.1?». Al realizar el test exacto de Fisher obtenemos la probabilidad de que los datos observados, Tabla 10.4.1, hayan surgido por azar, dado que los totales marginales (5 fallecimientos y 34 supervivientes, 10 que han recibido TMC y 29 que han recibido OMEC) son fijos. Para concretar esto un poco más, supongamos que la hipótesis nula es cierta y que se realiza un experimento en el que 10 bebés reciben TMC y 29 reciben OMEC. Además, supongamos que 5 de esos 39 bebés van a fallecer, independientemente del grupo en el que estén. Es decir, hay 5 bebés tan seriamente enfermos que ningún tratamiento podría salvarlos. ¿Cuál es la probabilidad de que 4 de ellos sean asignados al grupo de TMC?

Para obtener esta probabilidad, necesitamos determinar lo siguiente:

1. El número de formas de asignar exactamente 4 de los 5 bebés que van a fallecer al grupo de TMC.
2. El número de formas de asignar exactamente 6 de los 34 bebés que van a sobrevivir al grupo de TMC.
3. El número de formas de asignar 10 de los 39 bebés al grupo de TMC.

El producto de (1) y (2), dividido por (3), da la probabilidad en cuestión. ■

Combinaciones

En la Sección 3.6 presentamos la fórmula de la distribución binomial. Parte de su fórmula es el coeficiente nC_j (que en la Sección 3.6 denominamos coeficiente binomial). La cantidad nC_j es el número de formas en la que se pueden

escoger j objetos de un conjunto de n objetos. Por ejemplo, el número de formas en las que se puede escoger un grupo de 4 bebés entre un grupo de 5 bebés es ${}_5C_4$. El valor numérico de ${}_nC_j$ está dado por la fórmula 10.4.1:

$${}_nC_j = \frac{n!}{j!(n-j)!} \quad (10.4.1)$$

donde $n!$ («factorial de n ») se define para cualquier entero positivo como

$$n! = n(n-1)(n-2) \cdots (2)(1)$$

y $0! = 1$.

Por ejemplo, si $j = 1$ entonces tenemos ${}_nC_1 = \frac{n!}{1!(n-1)!} = n$, lo que tiene sentido: hay n formas de escoger un objeto en un conjunto de n objetos. Si $j = n$, entonces tenemos ${}_nC_n = \frac{n!}{n!0!} = 1$, ya que solo hay una forma de escoger en n objetos de un conjunto de tamaño n .

Ejemplo 10.4.2 OMEC

Podemos aplicar la fórmula 10.4.1 como sigue.

1. El número de formas de asignar 4 bebés al grupo TMC de entre los 5 que están destinados a fallecer es

$${}_5C_4 = \frac{5!}{4!1!} = 5.$$

2. El número de formas de asignar 6 bebés al grupo TMC de entre los 34 que van a sobrevivir es

$${}_{34}C_6 = \frac{34!}{6!28!} = 1.344.904.$$

3. El número de formas de asignar 10 bebés al grupo TMC de entre los 39 bebés totales es

$${}_{39}C_{10} = \frac{39!}{10!29!} = 635.745.396^*.$$

Por tanto, la probabilidad de obtener los mismos datos que en la Tabla 10.4.1, dado que los totales marginales son fijos, es $\frac{{}_5C_4 \times {}_{34}C_6}{{}_{39}C_{10}} = \frac{5 \times 1.344.904}{635.745.396} = 0,01058$. ■

Cuando se realiza el test exacto de Fisher de una hipótesis nula contra una alternativa direccional, necesitamos calcular las probabilidades de todas las tablas de datos (que tienen los mismos valores marginales que la tabla observada) que proporcionan una evidencia al menos tan fuerte contra H_0 , en la dirección predicha por H_A , como la tabla observada.

Ejemplo 10.4.3 OMEC

Previamente al experimento descrito en el Ejemplo 10.4.1, había evidencia que sugería que OMEC es mejor que TMC. Por tanto, es apropiada una hipótesis alternativa direccional:

$$H_A: \Pr\{\text{fallecimiento}|\text{OMEC}\} < \Pr\{\text{fallecimiento}|\text{TMC}\}$$

Los datos en la tabla observada, Tabla 10.4.1, dan fundamento a H_A . Hay otra posible tabla, que se presenta como Tabla 10.4.2, que tiene los mismos valores marginales que la Tabla 10.4.1, pero que es incluso más extrema en dar fundamento a H_A . Dado que 5 de los 39 bebés fallecieron y que 10 bebés se asignaron a TMC, el resultado más extremo posible que da fundamento a la hipótesis alternativa (que OMEC es mejor que TMC) es la tabla en la cual ninguno de los bebés que recibieron OMEC falleció y que los 5 fallecimientos ocurrieron en el grupo de TMC.

* Este ejemplo hace evidente que el uso de un computador o una calculadora gráfica es muy conveniente para realizar el test exacto de Fisher. En un procedimiento estadístico que casi nunca se utiliza sin la ayuda de la tecnología.

Tabla 10.4.2 Una tabla más extrema que podría haber resultado del experimento de OMEC

		Tratamiento		Total
		TMC	OMEC	
Resultado	Fallecen	5	0	5
	Sobreviven	5	29	34
Total		10	29	39

La probabilidad de que ocurra la Tabla 10.4.2, si H_0 es cierta, es $\frac{{}_5C_5 \times {}_{34}C_5}{{}_{39}C_{10}} = \frac{1 \times 278.256}{635.745.396} = 0,00044$. El P valor es la probabilidad de obtener datos al menos tan extremos como los observados, si H_0 es cierta. En este caso, el P valor es la probabilidad de obtener o bien los datos en la Tabla 10.4.1 o los de la Tabla 10.4.2, si H_0 es cierta. Por tanto, $P \text{ valor} = 0,01058 + 0,00044 = 0,01102$. Este P valor es bastante pequeño, por lo que el experimento proporciona una evidencia fuerte de que H_0 es falsa y que la OMEC es mejor que el TMC. ■

Comparación con el test chi-cuadrado

El test chi-cuadrado presentado en la Sección 10.2 se utiliza a menudo para analizar tablas de contingencia de 2×2 . Una ventaja del test chi-cuadrado es que se puede extender a tablas de 2×3 y otras tablas de dimensión mayor, como veremos en la Sección 10.6. El P valor del test chi-cuadrado se basa en la distribución chi-cuadrado, como su propio nombre implica. Se puede demostrar que, cuando el tamaño de la muestra se hace mayor, esta distribución proporciona una buena aproximación a la distribución muestral teórica del estadístico de contraste chi-cuadrado χ_s^2 . Si el tamaño de la muestra es pequeño, sin embargo, la aproximación puede ser pobre y el P valor del test chi-cuadrado puede ser engañoso.

El test exacto de Fisher se denomina «exacto» porque el P valor se determina de forma exacta, utilizando cálculos como el que se muestra en el Ejemplo 10.4.2, en vez de basarse en una aproximación asintótica. El Ejemplo 10.4.4 muestra cómo se comparan el test exacto y el test chi-cuadrado para los datos de OMEC.

Ejemplo 10.4.4 OMEC

Realizando un test chi-cuadrado sobre los datos experimentales de OMEC de la Tabla 10.4.1 se obtiene un estadístico de contraste de

$$\chi_s^2 = \frac{(4 - 1,28)^2}{1,28} + \frac{(1 - 3,72)^2}{3,72} + \frac{(6 - 8,72)^2}{8,72} + \frac{(28 - 25,28)^2}{25,28} = 8,89$$

El P valor (utilizando una alternativa direccional) es 0,0014. Este valor es bastante más pequeño que el P valor obtenido con el test exacto de 0,01102. ■

Test exacto y alternativas no direccionales

En general, la diferencia entre un contraste direccional y uno no direccional es que el P valor del contraste no direccional es dos veces el valor del contraste direccional (suponiendo que los datos se desvían con respecto a H_0 en la dirección especificada por H_A).

Para el test exacto de Fisher esto no es cierto. El P valor cuando H_A es no direccional no se obtiene simplemente multiplicando por dos el P valor del test direccional. En lugar de eso, un procedimiento generalmente aceptado es obtener las probabilidades de todas las tablas que son tan o menos probables que la tabla observada. Estas probabilidades se suman para obtener el P valor del test no direccional*. El Ejemplo 10.4.5 ilustra esta idea.

* No hay un acuerdo universal en este proceso. El P valor se puede calcular como la suma de las probabilidades de todas las tablas «extremas», pero hay varias formas de definir «extremo». Una alternativa al método presentado aquí es ordenar tablas de acuerdo al valor de χ_s^2 y contar como extremas aquellas tablas que tienen un valor mayor o igual que el valor de χ_s^2 obtenido de la tabla observada. Otro planteamiento es ordenar las tablas de acuerdo a $|p_1 - p_2|$. Estos métodos conducen algunas veces a obtener un P valor diferente al del método presentado aquí.

Ejemplo 10.4.5 Vacunas de la gripe

En una muestra aleatoria de estudiantes universitarios se encontró que 13 de ellos se habían puesto la vacuna de la gripe al comienzo del invierno y 28 no. De los 13 que se habían puesto la vacuna, 3 contrajeron la gripe durante el invierno. De los 28 que no se habían puesto la vacuna, 15 contrajeron la gripe²². Estos datos se muestran en la Tabla 10.4.3. Consideremos la hipótesis nula de que la probabilidad de contraer la gripe es la misma tanto si se ha vacunado como si no. La probabilidad de los datos en la Tabla 10.4.3, dado que los marginales son fijos, es $\frac{{}_{18}C_3 \times {}_{23}C_{10}}{{}_{41}C_{13}} = 0,05298$.

Tabla 10.4.3 Datos de vacunas de la gripe				
		Vacuna de la gripe		Total
		No vacuna		
¿Gripe?	Sí	15	3	18
	No	13	10	23
	Total	28	13	41

Una alternativa direccional natural sería que ponerse la vacuna de la gripe reduce la probabilidad de contraer la gripe. La Figura 10.4.1 muestra los datos obtenidos (de la Tabla 10.4.3), junto con las tablas de posibles resultados que dan un fundamento más fuerte a H_A . La probabilidad de cada tabla se muestra también en la Figura 10.4.1.

El P valor del test direccional es la suma de las probabilidades de estas tablas: $P \text{ valor} = 0,05298 + 0,01174 + 0,00138 + 0,00006 = 0,06616$.

Una alternativa no direccional plantea que la probabilidad de contraer la gripe depende de si se ha vacunado o no, pero no plantea si ponerse la vacuna aumenta o disminuye la probabilidad. (Algunas personas podrían contraer la gripe *debido* a la vacuna, por lo que podría ser creíble que el porcentaje total de personas que contraen la gripe sea mayor entre las que fueron vacunadas que entre las que no, ¡aunque los organismos públicos oficiales ciertamente esperan otra cosa!).

Tabla	Probabilidad				
<table> <tr><td>15</td><td>3</td></tr> <tr><td>13</td><td>10</td></tr> </table>	15	3	13	10	0,05298
15	3				
13	10				
<table> <tr><td>16</td><td>2</td></tr> <tr><td>12</td><td>11</td></tr> </table>	16	2	12	11	0,01174
16	2				
12	11				
<table> <tr><td>17</td><td>1</td></tr> <tr><td>11</td><td>12</td></tr> </table>	17	1	11	12	0,00138
17	1				
11	12				
<table> <tr><td>18</td><td>0</td></tr> <tr><td>10</td><td>13</td></tr> </table>	18	0	10	13	0,00006
18	0				
10	13				

Figura 10.4.1

Tabla	Probabilidad				
<table> <tr><td>5</td><td>13</td></tr> <tr><td>23</td><td>0</td></tr> </table>	5	13	23	0	0,00000
5	13				
23	0				
<table> <tr><td>6</td><td>12</td></tr> <tr><td>22</td><td>1</td></tr> </table>	6	12	22	1	0,00002
6	12				
22	1				
<table> <tr><td>7</td><td>11</td></tr> <tr><td>21</td><td>2</td></tr> </table>	7	11	21	2	0,00046
7	11				
21	2				
<table> <tr><td>8</td><td>10</td></tr> <tr><td>20</td><td>3</td></tr> </table>	8	10	20	3	0,00440
8	10				
20	3				
<table> <tr><td>9</td><td>9</td></tr> <tr><td>19</td><td>4</td></tr> </table>	9	9	19	4	0,02443
9	9				
19	4				
<table> <tr><td>10</td><td>8</td></tr> <tr><td>18</td><td>5</td></tr> </table>	10	8	18	5	0,08356
10	8				
18	5				

Figura 10.4.2

La Figura 10.4.2 muestra tablas de posibles resultados en los que el porcentaje de gripes es mayor entre aquellos que se pusieron la vacuna que entre aquellos que no. Se acompaña también la probabilidad de cada tabla. Las primeras cinco tablas tienen probabilidades menores que 0,05298, que es la probabilidad de los datos observados en la Tabla 10.4.3, pero la probabilidad de la sexta tabla es mayor que 0,05298. Por tanto, la contribución al P valor de este conjunto de tablas es la suma de las cinco primeras probabilidades: $0,00000 + 0,00002 + 0,00046 + 0,00440 + 0,02443 = 0,02931$. Sumando esto al P valor del test direccional, que era 0,06616, se obtiene el P valor del test no direccional: $P \text{ valor} = 0,06616 + 0,02931 = 0,09547$.

Como muestra el ejemplo, el cálculo del P valor para el test exacto de Fisher es bastante engorroso, particularmente cuando la alternativa es no direccional. Es altamente recomendable utilizar *software* estadístico para realizar este test. ■

Ejercicios 10.4.1-10.4.8

10.4.1 Considere la realización de un test exacto de Fisher con los siguientes datos ficticios de la tabla. Sea la hipótesis nula que el tratamiento y la respuesta son independientes y sea la alternativa la hipótesis direccional de que el tratamiento B es mejor que el tratamiento A. Escriba las tablas de posibles resultados que fundamentan más fuerte la hipótesis H_A .

		Tratamiento		Total
		A	B	
Resultado	Mueren	4	2	6
	Viven	10	14	24
	Total	14	16	30

10.4.2 Repita el Ejercicio 10.4.1 con la siguiente tabla de datos.

		Tratamiento		Total
		A	B	
Resultado	Mueren	5	3	8
	Viven	12	13	25
	Total	17	16	33

10.4.3 En una prueba clínica aleatorizada y doble ciego se proporcionó a 156 sujetos medicación antidepresiva para ayudarles a dejar de fumar. Un segundo grupo de 153 sujetos recibieron un placebo. El insomnio fue más común en el grupo de antidepresivos que en el grupo del placebo²³. El test exacto de Fisher sobre los datos de insomnio dio un P valor de 0,008. Interprete el P valor en el contexto de la prueba clínica.

10.4.4 (Ejercicio para computador) En una muestra aleatoria de 99 estudiantes de un conservatorio de música se encontró que 9 de las 48 mujeres muestreadas tenían «entonación perfecta» (la habilidad de identificar, sin error, el tono

de una nota musical), y solo 1 de los 51 hombres de la muestra tenía esa habilidad²⁴. Realice un test exacto de Fisher con la hipótesis nula de que tener entonación perfecta es independiente del sexo. Utilice una alternativa direccional y un valor de $\alpha = 0,05$. ¿Rechazaría H_0 ? ¿Por qué o por qué no?

10.4.5 Considere los datos del Ejercicio 10.4.4. Realice un test chi-cuadrado y compare los resultados de este test con los resultados del test exacto de Fisher.

10.4.6 (Ejercicio para computador) El factor de crecimiento pleiotrofina está asociado con la progresión del cáncer en humanos. En un intento de monitorizar el crecimiento de tumores, unos doctores midieron los niveles de pleiotrofina en suero en pacientes con cáncer de páncreas y en un grupo de pacientes de control. Obtuvieron que solo 2 de los 28 pacientes de control tenían niveles en suero por encima de dos desviaciones típicas respecto a la media del grupo de control, mientras que 20 de los 41 pacientes con cáncer tenían niveles en suero superiores²⁵. Utilice el test exacto de Fisher para determinar si es probable que una discrepancia tan grande (2 de 28 frente a 20 de 41) suceda por azar. Utilice una alternativa direccional y $\alpha = 0,05$.

10.4.7 (Ejercicio para computador) Un grupo de 225 hombres con hiperplasia prostática benigna se asignaron aleatoriamente para tomar extracto de palmito sierra o un placebo en un estudio doble ciego. Un año después, 45 de los 112 hombres en el grupo del palmito sierra (40 %) pensaban que estaban tomando palmito sierra, en comparación con 52 de los 113 hombres del grupo del placebo (46 %)²⁶. ¿Es esta diferencia consistente con una variación por azar? Realice un test exacto de Fisher utilizando una alternativa no direccional.

10.4.8 (Ejercicio para computador) En un experimento con sujetos que padecían esquizofrenia se comparó una «terapia personal» con una «terapia familiar». Sólo 2 de los 23 sujetos asignados a la terapia personal sufrieron recaídas psicóticas en el primer año del estudio, en comparación con 8 de los 24 sujetos asignados al grupo de terapia familiar²⁷. ¿Es esto evidencia suficiente para concluir, con un nivel de significación de 0,05, que los dos tipos de terapias no son igualmente efectivas? Realice un test exacto de Fisher utilizando una alternativa no direccional.

10.5 La tabla de contingencia $r \times k$

Las ideas de las Secciones 10.2 y 10.3 se pueden extender fácilmente a tablas de contingencia mayores de 2×2 . Consideraremos ahora una tabla de contingencia con r filas y k columnas, que se denomina **tabla de contingencia $r \times k$** . He aquí un ejemplo.

Ejemplo 10.5.1 Anidamiento de chorlitos

Unos ecólogos de la vida salvaje monitorizaron los hábitos de reproducción de chorlitos de montaña durante tres años y tomaron nota de donde anidaban. Encontraron 66 nidos en campos de agricultura (CA), 67 nidos en hábitats de hierba baja de perros de la pradera (PP) y 20 nidos en otros campos de hierba (H). La elección de los nidos varió a lo largo de los años para las 153 nidadas muestreadas. La Tabla 10.5.1 muestra los datos²⁸.

Tabla 10.5.1 Localizaciones de nidos de chorlitos a lo largo de los años				
Localización	Año			Total
	2004	2005	2006	
Campo agrícola (CA)	21	19	26	66
Hábitats de perros de la pradera (PP)	17	38	12	67
Campos de hierba (H)	5	6	9	20
Total	43	63	47	153

Para comparar las distribuciones de las tres localizaciones, podemos calcular los porcentajes por columnas, como se muestra en la Tabla 10.5.2. (Por ejemplo, en la muestra del 2004, $\frac{21}{43}$, o el 48,8 %, de los nidos estaban en campos agrícolas). Una inspección de la Tabla 10.5.2 muestra alguna diferencia clara entre las distribuciones de los tres porcentajes (columnas), siendo el hábitat de los perros de la pradera mucho más común en 2005 que en los otros años. ■

Tabla 10.5.2 Distribuciones porcentuales de nidos de chorlitos por año			
Localización	Año		
	2004	2005	2006
Campo agrícola (CA)	48,8	30,2	55,3
Hábitats de perros de la pradera (PP)	39,5	60,3	25,5
Campos de hierba (H)	11,6	9,5	19,1
Total	99,9*	100,0	99,9*
* Las sumas de los porcentajes de 2004 y de 2006 difieren del 100 % debido al redondeo.			

La Figura 10.5.1 es un diagrama de barras de los datos, que proporciona una impresión visual de las distribuciones.

El test t chi-cuadrado para la tabla $r \times k$

El objetivo del análisis estadístico de una tabla de contingencia $r \times k$ es investigar la relación entre la variable de las filas y la variable de las columnas. Esta investigación puede empezar con una inspección de los porcentajes por filas o por columnas, como en la Tabla 10.5.2. Un camino para el análisis posterior es preguntarse si las discrepancias en los

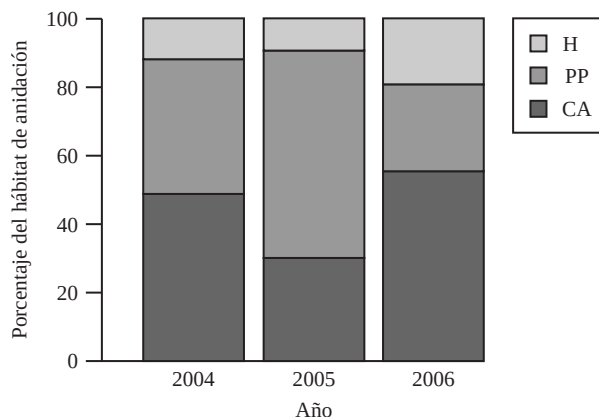


Figura 10.5.1 Diagrama de frecuencias relativas (porcentajes) apiladas de las localizaciones de los nidos de chorlitos por año

porcentajes son demasiado grandes como para ser explicadas por el error de muestreo. Esta cuestión se puede responder mediante un test chi-cuadrado. El estadístico chi-cuadrado se calcula mediante la fórmula habitual

$$\sum_{\text{todas las celdas}} \frac{(o_i - e_i)^2}{e_i}$$

dónde la suma se realiza sobre todas las $I = r \times k$ celdas de la tabla de contingencia, y las frecuencias esperadas (e) se calculan como

$$e = \frac{(\text{Total por filas}) \times (\text{Total por columnas})}{\text{Total completo}}$$

Este método de calcular los valores de e se puede justificar mediante una simple extensión del razonamiento dado en la Sección 10.2. Los valores críticos del test chi-cuadrado se obtienen a partir de la Tabla 9 con

$$gl = (r - 1)(k - 1)$$

El ejemplo siguiente ilustra el test chi-cuadrado.

Ejemplo 10.5.2 Anidamiento de chorlitos

Apliquemos el test chi-cuadrado a los datos sobre anidamiento de chorlitos del Ejemplo 10.5.1. La hipótesis nula es

H_0 : la distribución poblacional de las localizaciones de los nidos es la misma en los tres años.

Esta hipótesis se puede formular simbólicamente en notación de probabilidades condicionadas como sigue:

$$H_0: \begin{cases} \Pr\{CA|2004\} = \Pr\{CA|2005\} = \Pr\{CA|2006\} \\ \Pr\{PP|2004\} = \Pr\{PP|2005\} = \Pr\{PP|2006\} \\ \Pr\{H|2004\} = \Pr\{H|2005\} = \Pr\{H|2006\} \end{cases}$$

Nótese que los porcentajes de la Tabla 10.5.2 son las probabilidades condicionadas estimadas, es decir,

$$\hat{Pr}\{CA|2004\} = 0,488$$

$$\hat{Pr}\{CA|2005\} = 0,302$$

y así sucesivamente. Contrastaremos H_0 contra la hipótesis alternativa no direccional

H_A : la distribución poblacional de las localizaciones de los nidos no es la misma en los tres años.

La Tabla 10.5.3 muestra las frecuencias observadas y esperadas.

Tabla 10.5.3 Frecuencias observadas y esperadas de los nidos de chorlitos				
Localización	Año			Total
	2004	2005	2006	
Campo agrícola (CA)	21 (18,55)	19 (21,18)	26 (20,27)	66
Hábitats de perros de la pradera (PP)	17 (18,83)	38 (27,59)	12 (20,58)	67
Campos de hierba (H)	5 (5,62)	6 (8,24)	9 (6,14)	20
Total	43	63	47	153

Utilizando la Tabla 10.5.3 podemos calcular el estadístico de contraste como

$$\chi_s^2 = \frac{(21 - 18,55)^2}{18,55} + \frac{(19 - 21,18)^2}{21,18} + \dots + \frac{(9 - 6,14)^2}{6,14} = 14,09$$

Para esos datos, $r = 3$ y $k = 3$, por lo que

$$gl = (3 - 1)(3 - 1) = 4$$

Utilizando la Tabla 9 con $gl = 4$, obtenemos que $\chi_{4,0,01}^2 = 13,28$ y $\chi_{4,0,001}^2 = 18,47$, por lo que $0,001 < P$ valor $< 0,01$. Por tanto, el test chi-cuadrado muestra que hay evidencia significativa a favor de que las preferencias de localización de los nidos son diferentes a lo largo de los tres años. ■

Nótese que H_0 en el Ejemplo 10.5.2 es una hipótesis nula compuesta en el sentido definido en la Sección 9.4. Es decir, H_0 contiene más de una afirmación independiente. Esto siempre será verdad en tablas de contingencia de dimensiones mayores que 2×2 y, en consecuencia, para esas tablas la hipótesis alternativa del test chi-cuadrado será siempre no direccional y la conclusión, si H_0 se rechaza, será no direccional. Por tanto, el test chi-cuadrado a menudo no representa un análisis completo de una tabla de contingencia $r \times k$.

Dos contextos para las tablas de contingencia $r \times k$

Vimos en la Sección 10.3 que una tabla de contingencia 2×2 puede surgir en dos contextos diferentes. De forma similar, una tabla de contingencia $r \times k$ puede surgir en los dos contextos siguientes:

1. k muestras independientes. Una variable categórica observada con r categorías.
2. Una muestra. Dos variables categóricas observadas: una con k categorías y la otra con r categorías.

Como en el caso de la tabla 2×2 , el cálculo del estadístico chi-cuadrado es el mismo en ambos contextos, pero el planteamiento de las hipótesis y de las conclusiones puede ser diferente. El siguiente ejemplo ilustra el segundo contexto.

Ejemplo 10.5.3 Color de pelo y color de ojos

La Tabla 10.5.4 muestra la relación entre el color de pelo y el color de ojos de 6.800 hombres alemanes²⁹. (Es el mismo estudio que en el Ejemplo 10.3.2).

Utilizaremos un test chi-cuadrado para contrastar la hipótesis

H_0 : el color de pelo y el color de ojos son independientes.

Para los datos de la Tabla 10.5.4, se puede calcular $\chi_s^2 = 1,074$. Los grados de libertad del contraste son $gl = (3 - 1)(4 - 1) = 6$. Utilizando la Tabla 9 obtenemos que $\chi_{6,0,0001}^2 = 27,86$. Por tanto, H_0 es abrumadoramente rechazada y concluimos que hay una evidencia extremadamente fuerte a favor de que el color de pelo y el color de ojos están relacionados. ■

Tabla 10.5.4 Color de pelo y color de ojos

		Color de pelo			
		Marrón	Negro	Rubio	Rojo
Color de ojos	Marrones	438	288	115	16
	Grises o verdes	1.387	746	946	53
	Azules	807	189	1.768	47

Ejercicios 10.5.1-10.5.8

10.5.1 Unos pacientes con osteoartritis dolorosa de rodilla se asignaron aleatoriamente en un estudio clínico a uno de cinco tratamientos: glucosamina, condroitina, ambos, placebo o Celebrex, la terapia estándar. Un resultado que se apuntó fue si los pacientes experimentaban o no una mejoría sustancial en el dolor o presentaban recuperación funcional. Los datos se muestran en la tabla siguiente³⁰:

Tratamiento	Resultado satisfactorio		
	Tamaño de la muestra	Número	Porcentaje
Glucosamina	317	192	60,6
Condroitina	318	202	63,5
Ambos	317	208	65,6
Placebo	313	178	56,9
Celebrex	318	214	67,3

- (a) Utilice un test chi-cuadrado para comparar los porcentajes de éxitos con $\alpha = 0,05$. (El valor del estadístico chi-cuadrado es $\chi_s^2 = 9,29$).
- (b) Verifique el valor de χ_s^2 dado en el apartado (a).

10.5.2 En un estudio sobre poblaciones libres de la mosca *Drosophila subobscura*, unos investigadores colocaron trampas con cebo en dos lugares de un bosque y en un área abierta. El número de moscas macho y hembra atrapadas en un solo día se muestran en la tabla³¹.

	Lugar I del bosque	Lugar II del bosque	Área abierta
Machos	89	34	74
Hembras	31	20	136
Total	120	54	210

- (a) Utilice un test chi-cuadrado para comparar las proporciones de sexos en los tres sitios. Use $\alpha = 0,05$.
- (b) Construya una tabla que muestre los datos en un formato más legible, como el del Ejercicio 10.5.1.

10.5.3 En un estudio clásico sobre la úlcera péptica se determinaron los grupos sanguíneos de 1.655 pacientes con úlcera. La tabla que acompaña a este ejercicio muestra los da-

tos para estos pacientes y para un grupo elegido independientemente de 10.000 controles sanos de la misma ciudad³².

Grupo sanguíneo	Pacientes con úlcera	Controles
O	911	4.578
A	579	4.219
B	124	890
AB	41	313
Total	1.655	10.000

- (a) El valor del estadístico chi-cuadrado de la tabla de contingencia es $\chi_s^2 = 49,0$. Realice el test chi-cuadrado con $\alpha = 0,01$.
- (b) Construya una tabla que muestre la distribución de los porcentajes de los grupos sanguíneos en los pacientes y en los controles.
- (c) Verifique el valor de χ_s^2 dado en el apartado (a).

10.5.4 Las dos pinzas del bogavante (*Homarus americanus*) son idénticas en las etapas juveniles. Sin embargo, al llegar a adultos, las dos pinzas normalmente se han diferenciado en una pinza fuerte denominada «trituratora» y una pinza más delgada denominada «cortadora». En un estudio del proceso de diferenciación se criaron 26 animales juveniles en bandejas de plástico suave y se criaron 18 en bandejas que contenían trocitos de concha de ostra (que podían utilizar para ejercitar sus pinzas). Otros 23 animales se criaron en bandejas que contenían solo un trocito de cáscara de ostra. Las configuraciones de las pinzas de todos los animales cuando llegaron a adultos se resumen en la tabla³³.

Tratamiento	Configuración de las pinzas		
	Derecha trituradora. Izquierda cortadora	Derecha trituradora. Izquierda cortadora	Derecha trituradora. Izquierda cortadora
Trocitos de concha de ostra	8	9	1
Plástico suave	2	4	20
Un trocito de concha de ostra	7	9	7

- (a) El valor del estadístico chi-cuadrado de la tabla de contingencia para estos datos es $\chi_s^2 = 24,35$. Realice el test chi-cuadrado con $\alpha = 0,01$.
- (b) Verifique el valor de χ_s^2 dado en el apartado (a).
- (c) Construya una tabla que muestre la distribución de los porcentajes de las configuraciones de las pinzas en cada uno de los tres tratamientos.
- (d) Interprete la tabla del apartado (c): ¿De qué forma está la configuración de las pinzas relacionada con los tratamientos? (Por ejemplo, si se deseara un bogavante con dos pinzas cortadoras, ¿qué tratamiento escogería y por qué?)

10.5.5 Se realizó un experimento aleatorizado, doble ciego y controlado por placebo en el que unos pacientes con enfermedad de Alzheimer recibieron durante un año o bien extracto de Ginkgo biloba (EGb) o un placebo. Se midió el cambio de puntuación en la Escala de Evaluación de la Enfermedad de Alzheimer-subescala Cognitiva (EEEEA-Cog). Los resultados se muestran en la tabla³⁴. (Nota: Si el valor de EEEA-Cog disminuye, entonces es que el paciente mejora).

	Cambio en la puntuación de EEEA-Cog				
	-4 o mejor	-2 hasta -3	-1 hasta +1	+2 hasta +3	+4 o peor
EGb	22	18	12	7	16
Placebo	10	11	19	11	24

- (a) Utilice un test chi-cuadrado para comparar los porcentajes de prevalencia con $\alpha = 0,05$. (El valor del estadístico chi-cuadrado es $\chi_s^2 = 10,26$).
- (b) Verifique el valor de χ_s^2 dado en el apartado (a).

10.5.6 Los biólogos marinos piensan que el color de la banda de crecimiento más externa de las almejas tiende a estar relacionado con el momento del año en el que la almeja muere. Una bióloga realizó una pequeña investigación para ver si esto era cierto en la especie *Prptothaca staminea*. Recogió una muestra de 78 conchas de almejas de esta especie y realizó una clasificación cruzada de acuerdo con: (1) mes en el que la almeja murió y (2) color de la banda de crecimiento más externa. Los datos se muestran en la siguiente tabla³⁵.

	Color		
	Claro	Oscuro	Ilegible
Febrero	9	26	9
Marzo	6	25	3
Total	15	51	12

Utilice un test chi-cuadrado para comparar las distribuciones del color en los dos meses. Use $\alpha = 0,10$.

10.5.7 Un grupo de pacientes con problemas de comer en exceso se asignaron aleatoriamente o bien a tomar el medicamento experimental fluvoxamina o un placebo, en una prueba clínica doble ciego que duró nueve semanas. Al final de la prueba se clasificó la condición de cada paciente en una de cuatro categorías: sin respuesta, respuesta moderada, respuesta marcada o remisión. La tabla siguiente muestra una clasificación cruzada de los datos³⁶. ¿Hay evidencia estadísticamente significativa, al nivel de 0,10, para concluir que hay una asociación entre los grupos de tratamiento (fluvoxamina frente a placebo) y la condición?

	Sin respuesta	Respuesta moderada	Respuesta marcada	Remisión	Total
Fluvoxamina	15	7	3	15	40
Placebo	22	7	3	11	43
Total	37	14	6	26	

10.5.8 En una prueba clínica, unos pacientes con enfermedad de la arteria coronaria se asignaron aleatoriamente para recibir o bien angioplastia más terapia médica ($n = 1.149$) o únicamente terapia médica ($n = 1.138$). A lo largo de los años siguientes, 85 pacientes con angioplastia y 95 pacientes con solo terapia médica fallecieron, y la causa del fallecimiento se clasificó como cardíaca, otra, o desconocida. La tabla siguiente muestra una clasificación cruzada de los datos³⁷. ¿Existe evidencia estadísticamente significativa, al nivel de 0,10, para concluir que hay una asociación entre el grupo de tratamiento (angioplastia frente a terapia médica) y el resultado?

	Muerte por causa cardíaca	Muerte por otra causa	Muerte por causa desconocida	Vivos	Total
Angioplastia	23	45	17	1.064	1.149
Terapia médica	25	51	19	1.043	1.138
Total	48	96	36	2.107	2.287

10.6 Aplicabilidad de los métodos

En esta sección presentaremos orientaciones para decidir cuándo utilizar un test chi-cuadrado.

Condiciones de validez

Un test chi-cuadrado es válido bajo las siguientes condiciones:

1. *Condiciones de diseño.* En el test chi-cuadrado de una tabla de contingencia, debe ser apropiado ver los datos de una de las siguientes formas:
 - (a) Como dos o más muestras aleatorias independientes, observadas con respecto a una variable categórica.
 - (b) Como una muestra aleatoria, observada con respecto a dos variables categóricas.

Para cada tipo de test chi-cuadrado, las observaciones dentro de una muestra deben ser independientes entre sí.

2. *Condiciones del tamaño de la muestra.* El tamaño de la muestra debe ser suficientemente grande. Los valores críticos dados en la Tabla 9 son solo aproximadamente correctos para determinar el P valor asociado con χ^2_5 . Como una regla simple, la aproximación se considera adecuada si cada frecuencia esperada (e) es al menos igual a 5*. (Si las frecuencias esperadas son pequeñas y los datos forman una tabla de contingencia 2×2 , entonces el test exacto de Fisher podría ser apropiado, véase la Sección opcional 10.4).
3. *Forma de H_0 .* Una forma genérica de la hipótesis nula para el test chi-cuadrado de una tabla de contingencia se podría expresar como sigue:

H_0 : la variable de las filas y la variable de las columnas son independientes.

4. *Alcance de la inferencia.* Como con otros contrastes estadísticos, si los datos provienen de un experimento con asignación aleatoria de los tratamientos, como en el Ejemplo 10.1.1, entonces podemos extraer una inferencia causal. Si las unidades experimentales proceden de forma aleatoria de una población, entonces podemos extender la inferencia causal a esa población. Sin embargo, si los datos provienen de un estudio observacional, como en el Ejemplo 10.1.2, entonces un P valor pequeño solo permite inferir que la asociación observada no es debida al azar, pero no se pueden descartar otras explicaciones.

Verificación de las condiciones de diseño

Para verificar las condiciones de diseño, es necesario identificar una población de forma que los datos se puedan ver como una muestra aleatoria de la misma. Si los datos proceden de varias muestras [situación 1(a)], entonces se requiere que las muestras sean independientes entre sí. Si no se cumple esta restricción, el resultado puede ser una pérdida de potencia. Si el diseño incluye algún emparejamiento o igualación de unidades experimentales, entonces las muestras no serían independientes. En la Sección 10.8 se describe un método de análisis de muestras dependientes.

Como siempre, se debe excluir el sesgo en el procedimiento de muestreo. Además, los métodos chi-cuadrado no son apropiados cuando se utilizan esquemas complejos de muestreo aleatorio como muestreo por grupos o muestreo aleatorio estratificado. Finalmente, no debe haber dependencia ni estructuras jerárquicas en el diseño. Si no se cumple esta restricción, el resultado puede ser un aumento importante de la probabilidad del error de Tipo I (que es generalmente mucho más serio que una pérdida de potencia). Los ejemplos siguientes muestran la relevancia de comprobar la dependencia de las observaciones.

Ejemplo 10.6.1 Selección de alimento por larvas de insecto

En un estudio del comportamiento del gorgojo de la raíz del trébol de cuatro hojas *Sitona hospidulus* se liberaron 20 larvas en seis placas de Petri. Cada placa contenía raíces de alfalfa noduladas y no noduladas, dispuestas en un patrón simétrico. (Este experimento se describió de forma más completa en el Ejemplo 1.1.5). Después de 24 horas se apuntó la localización de cada larva, con los resultados que se muestran en la Tabla 10.6.1³⁸.

* En una tabla de contingencia $r \times k$ con más de 2 filas y columnas, la aproximación es adecuada si la frecuencia esperada promedio es al menos 5, incluso si en algunas celdas es menor.

Tabla 10.6.1 Elección de alimento por larvas <i>Sitona</i>			
Placa	Número de larvas		
	Raíces noduladas	Raíces no noduladas	Otras (muertas, pérdidas, etc.)
1	5	3	12
2	9	1	10
3	6	3	11
4	7	1	12
5	5	1	14
6	14	3	3
Total	46	12	62

Supongamos que se propone el siguiente análisis. Un total de 58 larvas realizaron una elección. Las frecuencias observadas de elección de raíces noduladas y no noduladas fueron 46 y 12, y las frecuencias esperadas correspondientes (suponiendo una elección aleatoria) serían de 29 y 29. Estos datos permiten calcular $\chi_s^2 = 19,93$, de donde (utilizando la alternativa direccional) obtenemos de la Tabla 9 que P valor $< 0,00005$. La validez de este análisis propuesto es altamente dudosa porque depende del supuesto de que todas las observaciones de una placa dada son independientes entre sí. Este supuesto sería falso si (como es biológicamente creíble), las larvas tienden a seguirse entre ellas en su búsqueda de alimento.

¿Como deberían ser analizados los datos entonces? Un planteamiento es realizar la suposición razonable de que las observaciones de una placa son independientes de las de otra placa. Bajo esta suposición se podría utilizar un análisis pareado de las seis placas ($n_d = 6$). Un test t de muestras pareadas permite calcular un P valor $\approx 0,005$ y un test de signos da un P valor $\approx 0,02$. Nótese que el supuesto cuestionable de independencia dentro de las placas produjo un P valor que era mucho más pequeño. ■

Ejemplo 10.6.2 Polinización de flores

Se realizó un estudio para determinar la significación adaptativa del color de las flores en la gilia escarlata (*Ipomopsis aggregata*). Se escogieron seis plantas con flores rojas y seis plantas con flores blancas para su observación en condiciones de campo. Se permitió que los colibríes visitaran las flores, pero se evitó que el otro polinizador principal, una palomilla, visitara las plantas por la noche. La Tabla 10.6.2 muestra, para cada planta, el número total de flores al final de la estación y el número de ellas que habían producido fruto.³⁹

Tabla 10.6.2 Frutos en flores de la gilia escarlata						
	Plantas con flores rojas			Plantas con flores blancas		
	Número de flores	Número de frutos	Porcentaje de frutos	Número de flores	Número de frutos	Porcentaje de frutos
	140	26	19	125	21	17
	116	11	9	134	17	13
	34	0	0	273	81	30
	79	9	11	146	38	26
	185	28	15	103	17	17
	106	11	10	82	24	29
Suma total	660	85		863	198	

La cuestión de interés es si el porcentaje de frutos es diferente en las plantas con flores rojas que en las plantas con flores blancas. Supongamos que esta cuestión se aborda considerando la flor individual como unidad observacional. Entonces, los datos se podrían disponer en el formato de la tabla de contingencia que se presenta en la Tabla 10.6.3.

Tabla 10.6.3 Frutos en flores de la gilia escarlata			
		Color de la flor	
		Rojo	Blanco
Fruto seco	Sí	85	198
	No	575	665
Total		660	863
Porcentaje de frutos		13	23

La Tabla 10.6.3 permite calcular $\chi_s^2 = 25,0$, de donde la Tabla 9 da un P valor $< 0,0001$. Sin embargo, este análisis no es correcto, porque las observaciones en flores de la misma planta no son independientes entre sí. Son dependientes porque el polinizador (el colibrí) tiende a visitar las flores en grupos, y quizá también porque las flores de la misma planta están relacionadas fisiológicamente y genéticamente. El test chi-cuadrado queda invalidado por la estructura jerárquica de los datos.

Un planteamiento mejor sería tratar toda una planta como la unidad observacional. Por ejemplo, se podría tomar la columna «Porcentaje de frutos» de la Tabla 10.6.2 como las observaciones básicas. Aplicando un test t a los valores se obtiene $t_s = 2,88$ (con $0,01 < P$ valor $< 0,02$) y aplicando un test de Wilcoxon-Mann-Whitney resulta $U_s = 32$ (con $0,02 < P$ valor $< 0,05$). Por tanto, el P valor del análisis chi-cuadrado inapropiado es mucho más pequeño. ■

Consideraciones sobre la potencia

En muchos estudios el test chi-cuadrado es válido pero no es tan potente como un test más apropiado. Concretamente, consideremos una situación en la cual las filas o las columnas (o ambas) de la tabla de contingencia corresponden a una variable categórica *ordenable* con más de dos categorías. A continuación se presenta un ejemplo.

Ejemplo 10.6.3 Fisioterapia

Se realizó una prueba clínica aleatorizada para determinar si la adición de sesiones de fisioterapia en sábado (el «tratamiento») a las sesiones habituales de lunes a viernes (el «control») sería beneficioso para los pacientes que asistían a rehabilitación en un hospital. Una medida de resultados fue el destino de un paciente después de terminar la rehabilitación, con las categorías de: casa, cuidados residenciales de bajo nivel (CRBN), cuidados residenciales de alto nivel (CRAN) o agudización con transferencia a un hospital (ATH). Los resultados se muestran en la Tabla 10.6.4⁴⁰.

Tabla 10.6.4 Destino de pacientes de fisioterapia			
		Grupo	
		Tratamiento	Control
Destino	Casa	107	103
	CRBN	10	15
	CRAN	6	1
	ATH	7	13
	Total	130	132

Sería válido un test chi-cuadrado de la tabla de contingencia para comparar los tratamientos y el control, pero el test perdería potencia debido a que no utiliza la información contenida en el *orden* de las categorías de los destinos (ir a casa es preferible a CRBN, que a su vez es preferible a CRAN, que a su vez es preferible a ATH). Una debilidad relacionada del test chi-cuadrado es que, incluso si H_0 se rechaza, el test no produce una conclusión direccional como: «el tratamiento produce destinos mejores que el control».

Existen métodos para analizar tablas de contingencia con variables de filas y/o de columnas ordenables. Estos métodos, sin embargo, están más allá del alcance de este libro.

Ejercicios 10.6.1-10.6.3

10.6.1 En referencia a los datos de quimioterapia del Ejercicio 10.2.10, ¿son los tamaños de las muestras suficientemente grandes para la validez aproximada del test chi-cuadrado?

10.6.2 En un estudio sobre las influencias prenatales en la susceptibilidad a los ataques en ratones, se asignaron aleatoriamente hembras embarazadas a un grupo de control o un grupo de «tratamiento». Se suministró a los ratones tratados inyecciones inocuas tres veces durante la gestación, mientras que no se tocó a los ratones de control. Se comprobó la susceptibilidad de la descendencia a ataques producidos por un ruido fuerte. Los investigadores advirtieron que la respuesta variaba considerablemente de camada a camada. La tabla que acompaña a este ejercicio resume los resultados⁴¹.

Si se analizan estos datos mediante una tabla de contingencia de 2×3 , el estadístico chi-cuadrado es $\chi_s^2 = 8,45$ y la Tabla 9 permite obtener $0,01 < P \text{ valor} < 0,02$. ¿Es este análisis apropiado para este experimento? Explique su respuesta. (*Sugerencia:* ¿cumple el diseño las condiciones de validez del test chi-cuadrado?).

			Respuesta a un ruido fuerte		
Tratamiento	Número de camadas	Número de ratones	Sin respuesta	Carrera salvaje	Ataque
Tratados	19	104	23	10	71
Control	20	120	47	13	60

10.6.3 En el control de la diabetes es importante conocer cómo cambian los niveles de glucosa en sangre después de

tomar varios alimentos. Diez voluntarios participaron en un estudio para comparar los efectos de dos alimentos: azúcar y almidón. Se extrajo un espécimen de sangre antes de que cada voluntario consumiera una cantidad medida de alimento. Después se tomaron especímenes de sangre adicionales 11 veces durante las siguientes cuatro horas. Los voluntarios repitieron la prueba completa en otra ocasión con el otro alimento. Era de especial interés la disminución de los niveles de glucosa por debajo del valor inicial. La tabla que acompaña a este ejercicio muestra estos valores⁴².

Alimento	Número de valores menores que el valor inicial	Número total de observaciones
Azúcar	26	110
Almidón	14	110

Suponga que se analizan los datos dados como una tabla de contingencia. El estadístico de contraste sería

$$\chi_s^2 = \frac{(26 - 20)^2}{20} + \frac{(14 - 20)^2}{20} + \frac{(84 - 90)^2}{90} + \frac{(96 - 90)^2}{90} = 4,40$$

Con $\alpha = 0,05$ rechazaríamos H_0 y obtendríamos que hay evidencia suficiente para concluir que los valores de glucosa en sangre por debajo del valor inicial aparecen más a menudo después de la ingestión de azúcar que después de la ingestión de almidón. Este análisis contiene dos errores. ¿Cuáles son? (*Sugerencia:* ¿se cumplen las condiciones para la validez del test?).

10.7 Intervalo de confianza para diferencia de probabilidades

El test chi-cuadrado de una tabla de contingencia 2×2 responde solo a una cuestión limitada: ¿Difieren lo suficiente las probabilidades estimadas (denominadas $\hat{p}_1$ y $\hat{p}_2$) como para concluir que las verdaderas probabilidades (denominadas p_1 y p_2) no son iguales? Un procedimiento complementario de análisis es construir un intervalo de confianza para la magnitud de la diferencia ($p_1 - p_2$).

Al explicar la construcción del intervalo de confianza para una proporción, p , en la Sección 9.2, definimos un estimador $\tilde{p}$, basado en la idea de «añadir dos éxitos y dos fracasos a los datos». Al hacer este ajuste se conseguía un procedimiento de obtención de intervalos de confianza con buenas propiedades de cobertura. De la misma forma, al construir un

intervalo de confianza para una diferencia de dos proporciones, definiremos nuevos estimadores que se basan en añadir 1 observación a cada celda de la tabla (de forma que en *total* se añaden dos éxitos y dos fracasos a los datos).

Consideremos una tabla de contingencia de 2×2 que se puede ver como una comparación de dos muestras, de tamaños n_1 y n_2 , con respecto a una variable de respuesta dicotómica. Sea la tabla de 2×2 :

Muestra 1	Muestra 2
y_1	y_2
$n_1 - y_1$	$n_2 - y_2$
n_1	n_2

Definimos

$$\tilde{p}_1 = \frac{y_1 + 1}{n_1 + 2}$$

y

$$\tilde{p}_2 = \frac{y_2 + 1}{n_2 + 2}$$

Utilizaremos la diferencia de los nuevos valores, $(\tilde{p}_1 - \tilde{p}_2)$ para construir un intervalo de confianza para $(p_1 - p_2)$. Como todos los valores calculados a partir de muestras, la cantidad $(\tilde{p}_1 - \tilde{p}_2)$ está sujeta al error de muestreo. La magnitud del error de muestreo se puede expresar mediante el error típico de $(\tilde{p}_1 - \tilde{p}_2)$, que se calcula mediante la siguiente fórmula:

$$ET_{(\tilde{p}_1 - \tilde{p}_2)} = \sqrt{\frac{\tilde{p}_1(1 - \tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1 - \tilde{p}_2)}{n_2 + 2}}$$

Nótese que $ET_{(\tilde{p}_1 - \tilde{p}_2)}$ es análogo a $ET_{(\bar{y}_1 - \bar{y}_2)}$, descrito en la Sección 6.6.

Un intervalo de confianza aproximado se puede basar en $ET_{(\tilde{p}_1 - \tilde{p}_2)}$. Por ejemplo, un intervalo de confianza del 95 % es

$$(\tilde{p}_1 - \tilde{p}_2) \pm (1,96) ET_{(\tilde{p}_1 - \tilde{p}_2)}$$

Los intervalos de confianza contruidos de esta forma tienen buenas propiedades de cobertura (es decir, aproximadamente el 95 % de todos los intervalos de confianza del 95 % contendrán a la verdadera diferencia $(p_1 - p_2)$ para tamaños muestrales cualesquiera n_1 y n_2 ⁴³. El ejemplo siguiente ilustra la construcción del intervalo de confianza*.

Ejemplo 10.7.1 Migrañas

En los datos sobre migrañas del Ejemplo 10.1.1, los tamaños muestrales eran $n_1 = 49$ y $n_2 = 26$, y las probabilidades estimadas de una reducción sustancial de migrañas eran

$$\tilde{p}_1 = \frac{42}{51} = 0,824$$

$$\tilde{p}_2 = \frac{16}{28} = 0,571$$

* En la Sección 9.3 presentamos una versión general de la idea de «sumar dos éxitos y dos fracasos», en la que la fórmula de $\tilde{p}$ dependía del nivel de confianza (95 %, 90 %, etc.). Al construir el intervalo de confianza para la diferencia de proporciones, las propiedades de cobertura del intervalo son mejores cuando se añade 1 a cada celda de la tabla 2×2 , independientemente del nivel de confianza utilizado⁴⁴.

La diferencia entre ellas es

$$\begin{aligned}\tilde{p}_1 - \tilde{p}_2 &= 0,824 - 0,571 \\ &= 0,253 \\ &\approx 0,25\end{aligned}$$

Por tanto, estimamos que la cirugía real aumenta la probabilidad de reducción sustancial de migrañas en 0,25, comparada con la cirugía falsa de imitación. Para establecer límites de confianza para esta estimación, calculamos el error típico como

$$ET_{(\tilde{p}_1 - \tilde{p}_2)} = \sqrt{\frac{0,824(0,176)}{51} + \frac{0,571(0,429)}{28}} = 0,1077$$

El intervalo de confianza del 95 % es

$$\begin{aligned}0,253 \pm (1,96)(0,1077) \\ 0,253 \pm 0,211 \\ 0,042 < p_1 - p_2 < 0,464\end{aligned}$$

Tenemos una confianza del 95 % en que la probabilidad de una reducción sustancial en las migrañas sea entre 0,042 y 0,464 más alta con la cirugía real que con la cirugía falsa de imitación. ■

Relación con el test El test chi-cuadrado para una tabla de contingencia de 2×2 (Sección 10.2) es aproximadamente, pero no exactamente, igual a comprobar si un intervalo de confianza para $(p_1 - p_2)$ incluye al cero. [Recuérdese de la Sección 7.3 que hay una equivalencia exacta entre el test t y el intervalo de confianza para $(\mu_1 - \mu_2)$].

Ejercicios 10.7.1-10.7.6

10.7.1 En una prueba clínica doble ciego, una serie de pacientes ancianos que habían sufrido fracturas de cadera se asignaron aleatoriamente para recibir o bien un placebo ($n = 1.062$) o ácido zolendrónico ($n = 1.065$). Durante la prueba, 139 pacientes que tomaron placebo y 92 pacientes que tomaron ácido zolendrónico tuvieron nuevas fracturas⁴⁵. Sean p_1 y p_2 las probabilidades de fractura con placebo y con ácido zolendrónico, respectivamente. Construya un intervalo de confianza del 95 % para $(p_1 - p_2)$.

10.7.2 En referencia a los datos sobre tumores de hígado del Ejercicio 10.2.9:

- Construya un intervalo de confianza del 95 % para $(\Pr\{\text{tumor de hígado}|\text{sin gérmenes}\} - \Pr\{\text{tumor de hígado}|E. coli\})$.
- Interprete el intervalo de confianza del apartado (a). Es decir, explique lo que indica el intervalo sobre las probabilidades de tumor.

10.7.3 En mujeres embarazadas de gemelos, generalmente se prescribe reposo en las últimas fases del embarazo para reducir el riesgo de parto prematuro. Para comprobar el valor práctico de esta medida, se asignaron aleatoriamente 212 mujeres embarazadas de gemelos a un grupo de reposo en cama o a un grupo de control. La tabla siguiente muestra la incidencia de partos prematuros (menos de 37 semanas de gestación)⁴⁶.

	Reposo en cama	Controles
Número de partos prematuros	32	20
Número de mujeres	105	107

Construya un intervalo de confianza del 95 % para $(\Pr\{\text{parto prematuro}|\text{cama}\} - \Pr\{\text{parto prematuro}|\text{control}\})$. ¿Sugiere este intervalo de confianza que el reposo en cama es beneficioso?

10.7.4 En referencia al Ejercicio 10.7.3, el número de niños con bajo peso al nacer (2.500 g o menos) en los partos de esas mujeres se muestran en la tabla siguiente.

	Reposo en cama	Controles
Número de partos prematuros	76	92
Número de mujeres	210	214

Sean p_1 y p_2 las probabilidades de un bebé con bajo peso al nacer en las dos condiciones. Explique por qué la información anterior no es suficiente para construir un intervalo de confianza para $(p_1 - p_2)$.

10.7.5 En referencia a los datos de grupo sanguíneo del Ejercicio 10.5.3, sean p_1 y p_2 las probabilidades de sangre de

tipo O en la población de pacientes y en la población de control, respectivamente.

- Construya un intervalo de confianza del 95 % para $(p_1 - p_2)$.
- Interprete el intervalo de confianza del apartado (a). Es decir, explique lo que indica el intervalo acerca de las diferencias de las probabilidades de sangre de tipo O.

10.7.6 En un experimento para tratar pacientes con «trastorno de ansiedad generalizado» se suministró el medicamento hidroxizina a 71 pacientes y 30 de ellos mejoraron. Se suministró un placebo a un grupo de 70 pacientes y 20 de ellos mejoraron⁴⁷. Sean p_1 y p_2 las probabilidades de mejora utilizando hidroxizina y utilizando el placebo, respectivamente. Construya un intervalo de confianza del 95 % para $(p_1 - p_2)$.

10.8 Datos pareados y tablas de 2×2 (opcional)

En el Capítulo 8 consideramos datos pareados cuando la variable de respuesta era continua. En esta sección consideraremos el análisis de datos categóricos pareados.

Ejemplo 10.8.1 Transmisión del VIH a niños

Se realizó un estudio para determinar el riesgo que tiene una mujer de transmitir el VIH a su hijo durante el embarazo. En una muestra de 114 mujeres infectadas con el VIH que tuvieron dos niños se encontró que 19 de los 114 hermanos mayores estaban infectados y que 20 de los hermanos menores estaban infectados⁴⁸. Estos datos se muestran en la Tabla 10.8.1.

Tabla 10.8.1 Datos de infección por VIH			
		Hermano mayor	Hermano menor
¿VIH?	Sí	19	20
	No	95	94
	Total	114	114

A primera vista, podría parecer que se podría utilizar un test chi-cuadrado normal para contrastar la hipótesis nula de que la probabilidad de infección por VIH es la misma para los hermanos mayores que para los hermanos menores. Sin embargo, como indicamos en la Sección 10.6, para que el test chi-cuadrado sea válido las dos muestras (los 114 hermanos mayores y los 114 hermanos menores) deben ser independientes entre sí. En este caso las muestras son claramente dependientes. De hecho, se trata de datos emparejados, con una familia que genera la pareja (hermano mayor, hermano menor).

La Tabla 10.8.2 presenta los datos en un formato diferente. Este formato ayuda a centrar la atención en la parte relevante de los datos*.

Tabla 10.8.2 Datos de infección por VIH mostrados por parejas			
		Hermano menor ¿VIH?	
		Sí	No
Hermano mayor ¿VIH?	Sí	2	17
	No	18	77

En la Tabla 10.8.2 podemos ver que hay 79 parejas en las que ambos hermanos tienen el mismo estado respecto al VIH: dos son parejas «sí/sí» y 77 son parejas «no/no». Esas 79 parejas, que se denominan **parejas concordantes**, no ayudan a determinar si la infección es más probable en hermanos menores que en hermanos mayores. Las restan-

* Nótese que la Tabla 10.8.2 no se puede obtener a partir de la Tabla 10.8.1.

tes 35 parejas (17 parejas «sí/no» y 18 parejas «no/sí») sí proporcionan información sobre las probabilidades relativas de infección por VIH en hermanos mayores y menores. Esas parejas se denominan **parejas discordantes**. Nuestro análisis se centrará en esas 35 parejas.

Si la probabilidad de infección por VIH es la misma en hermanos mayores que en hermanos menores, entonces los dos tipos de parejas («sí/no» y «no/sí») son igualmente probables. Por tanto, la hipótesis nula

H_0 : la probabilidad de infección por VIH es la misma en hermanos mayores que en hermanos menores es equivalente a

$$H_0: \text{entre parejas discordantes, } \Pr(\text{«sí/no»}) = \Pr(\text{«no/sí»}) = \frac{1}{2}. \quad \blacksquare$$

El test de McNemar

La hipótesis de que es igualmente probable que las parejas discordantes sean «sí/no» que «no/sí» se puede comprobar con el test chi-cuadrado de bondad de ajuste desarrollado en la Sección 9.4. La aplicación del test chi-cuadrado se conoce con el nombre de **test de McNemar** y tiene una forma particularmente simple*. Sea n_{11} el número de parejas «sí/sí», n_{12} el número de parejas «sí/no», n_{21} el número de parejas «no/sí» y n_{22} el número de parejas «no/no», como se muestra en la Tabla 10.8.3. Si H_0 es cierta, el número esperado de parejas «sí/no» es $\frac{n_{12} + n_{21}}{2}$, que coincide con el número esperado de

parejas «no/sí». Por tanto, el estadístico de contraste es

$$\chi_s^2 = \frac{\left(n_{12} - \frac{(n_{12} + n_{21})}{2}\right)^2}{\frac{(n_{12} + n_{21})}{2}} + \frac{\left(n_{21} - \frac{(n_{12} + n_{21})}{2}\right)^2}{\frac{(n_{12} + n_{21})}{2}}$$

que se puede simplificar a

$$\chi_s^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

La distribución de χ_s^2 bajo la hipótesis nula es aproximadamente una distribución χ^2 con un grado de libertad.

Tabla 10.8.3 Tabla general de datos de proporciones emparejados		
	Sí	No
Sí	n_{11}	n_{12}
No	n_{21}	n_{22}

Ejemplo 10.8.2 Transmisión de VIH a niños

Para los datos dados en el Ejemplo 10.8.1, $n_{12} = 17$ y $n_{21} = 18$. Por tanto,

$$\chi_s^2 = \frac{(17 - 18)^2}{17 + 18} = 0,0286$$

Utilizando la Tabla 9, vemos que el P valor es mayor que 0,20. (Utilizando un computador se obtiene un P valor de 0,87). Los datos son muy consistentes con la hipótesis nula de que la probabilidad de infección por VIH es la misma para hermanos mayores que para hermanos menores. ■

* La hipótesis nula contrastada en el test de McNemar también se puede contrastar utilizando la distribución binomial. La hipótesis nula establece que entre las parejas discordantes, $\Pr(\text{«sí/no»}) = \Pr(\text{«no/sí»}) = 1/2$. Por tanto, bajo la hipótesis nula, el número de parejas «sí/no» tiene una distribución binomial con n = número de parejas discordantes y $p = 0,5$.

Ejercicios 10.8.1-10.8.4

10.8.1 Como parte de un estudio sobre factores de riesgo de hemorragia cerebral, se entrevistó a 155 mujeres (casos) que habían experimentado dicho tipo de hemorragia. En cada caso se eligió un control que no había tenido hemorragia. El control se igualó por proximidad de residencia, edad y raza. Se preguntó a cada mujer si utilizaba anticonceptivos orales. Los datos de las 155 parejas se muestran en la tabla. «Sí» y «No» se refieren al uso de anticonceptivos orales.

		Caso	
		No	Sí
Control	No	107	30
	Sí	13	5

Para contrastar la asociación entre el uso de anticonceptivos orales y la hemorragia cerebral, considere solo las 43 parejas discordantes (parejas que contestaron de forma diferente), y compruebe la hipótesis de que es igualmente probable que una pareja discordante sea «sí/no» que «no/sí». Utilice el test de McNemar para contrastar la hipótesis de que tener una hemorragia cerebral es independiente de usar anticonceptivos orales, contra una alternativa no direccional con $\alpha = 0,05$.

10.8.2 El Ejemplo 10.8.1 se refería a una muestra de mujeres infectadas con el VIH que tuvieron dos hijos. Uno de los resultados que se estudió fue si el tiempo de gestación del niño fue inferior a 38 semanas. Se recogió esta información en 106 familias. Los datos de esta variable se muestran en la tabla siguiente. Analice estos datos utilizando el test de McNemar. Use una alternativa no direccional y $\alpha = 0,10$.

		¿Hermano menor < 38 semanas?	
		Sí	No
¿Hermano mayor < 38 semanas?	Sí	26	5
	No	21	54

10.8.3 En un estudio sobre 85 pacientes con la enfermedad de Hodgkin se encontró que a 41 de ellos se les habían extirpado las amígdalas. Cada pariente se emparejó con un hermano del mismo sexo. Sólo a 33 de los hermanos se les había practicado una amigdalectomía. Los datos se muestran en la tabla siguiente⁵⁰. Utilice el test de McNemar para contrastar la hipótesis de que las parejas «sí/no» y «no/sí» son equiprobables. Hay investigaciones previas que han sugerido que la extirpación de las amígdalas está asociado con un riesgo mayor de padecer la enfermedad de Hodgkin. Por tanto, utilice una alternativa direccional. Use $\alpha = 0,05$.

		¿Amigdalectomía al hermano?	
		Sí	No
Amigdalectomía al paciente de Hodgkin	Sí	26	15
	No	7	37

10.8.4 En un estudio sobre el comportamiento sexual del *Gryllus campestris* se situaron sobre plexiglás parejas de grillos hembra junto con un grillo macho. En 54 casos las hembras lucharon, y después en 42 casos la hembra victoriosa copuló con el macho, en 8 casos la hembra derrotada copuló con el macho y en 4 casos no hubo copulación. Los datos se resumen en la tabla siguiente⁵¹. Utilice el test de McNemar para contrastar la hipótesis de que las hembras ganadoras y las perdedoras tienen la misma probabilidad de copular. Emplee la alternativa direccional apropiada y $\alpha = 0,05$.

		Ganadoras	
		Sí	No
Cópula	Sí	0	8
	No	42	4

10.9 Riesgo relativo y razón de ventajas (opcional)

Es bastante común contrastar la hipótesis de que dos proporciones poblacionales, p_1 y p_2 , son iguales. A menudo se utiliza para este propósito un test chi-cuadrado basado en una tabla de 2×2 . El intervalo de confianza para $(p_1 - p_2)$ proporciona información sobre la magnitud de la diferencia entre p_1 y p_2 . En esta sección consideraremos otras dos medidas de dependencia: el riesgo relativo y la razón de ventajas.

Riesgo relativo

Algunas veces los investigadores prefieren comparar probabilidades en función de su *razón*, y no de su diferencia. Cuando el suceso resultado es perjudicial (como tener un ataque cardíaco o tener cáncer) la razón de probabilidades se denomina **riesgo relativo**, o razón de riesgo. El riesgo relativo se define como p_1/p_2 . Esta medida es ampliamente utilizada en estudios sobre la salud humana. Se presenta a continuación un ejemplo.

Ejemplo 10.9.1 Fumar y cáncer de pulmón

Se realizó un seguimiento durante muchos años de las historias clínicas de 11.900 hombres de mediana edad. Durante el estudio, 126 hombres contrajeron cáncer de pulmón. Entre ellos había 89 fumadores y 37 ex fumadores. La Tabla 10.9.1 muestra los datos⁵².

Tabla 10.9.1 Incidencia del cáncer de pulmón en fumadores y ex fumadores			
		Historia del fumador	
		Fumador	Ex fumador
¿Cáncer de pulmón?	Sí	89	37
	No	6.063	5.711
	Total	6.152	5.748

Las probabilidades de interés principal son las probabilidades condicionales por columnas:

$$p_1 = \Pr\{\text{cáncer de pulmón}|\text{fumador}\}$$

$$p_2 = \Pr\{\text{cáncer de pulmón}|\text{ex fumador}\}$$

Sus estimaciones a partir de los datos son

$$\hat{p}_1 = \frac{89}{6.152} = 0,01447 \approx 0,014$$

$$\hat{p}_2 = \frac{37}{5.748} = 0,00644 \approx 0,006$$

El riesgo relativo estimado es

$$\frac{\hat{p}_1}{\hat{p}_2} = \frac{0,01447}{0,00644} = 2,247 \approx 2,2$$

Por tanto, estimamos que el riesgo (es decir, la probabilidad condicional) de contraer cáncer de pulmón es aproximadamente 2,2 veces mayor en fumadores que en ex-fumadores. (Por supuesto, dado que se trata de un estudio observacional, no estaría justificado concluir que fumar *causa* cáncer de pulmón). ■

La razón de ventajas

Otra forma de comparar dos probabilidades es en términos de **ventajas**. La ventaja de un suceso E se define como la razón de las probabilidades de que E ocurra frente a que E no ocurra:

$$\text{ventaja de } E = \frac{\Pr\{E\}}{1 - \Pr\{E\}}$$

Por ejemplo, si la probabilidad de un suceso es de $1/4$, entonces la ventaja del suceso es $\frac{1/4}{1/3} = 1/3$ o $1:3$. Como otro ejemplo, si la probabilidad de un suceso es $1/2$, entonces la ventaja del suceso son $\frac{1/2}{1/2} = 1$ o $1:1$.

La **razón de ventajas** es simplemente el cociente de la ventaja de un suceso bajo dos condiciones. Concretamente, supongamos que p_1 y p_2 son las probabilidades condicionales de un suceso bajo dos condiciones diferentes. Entonces la razón de ventajas, que denominaremos θ («theta») se define como sigue:

$$\theta = \frac{\frac{p_1}{1 - p_1}}{\frac{p_2}{1 - p_2}}$$

Si las probabilidades estimadas $\hat{p}_1$ y $\hat{p}_2$ se calculan mediante una tabla de contingencia de 2×2 , la correspondiente razón de ventajas estimada, denominada $\hat{\theta}$, se calcula como

$$\hat{\theta} = \frac{\frac{\hat{p}_1}{1 - \hat{p}_1}}{\frac{\hat{p}_2}{1 - \hat{p}_2}}$$

Lo ilustraremos con un ejemplo.

Ejemplo 10.9.2 Fumar y cáncer de pulmón

A partir de los datos del Ejemplo 10.9.1, estimamos la ventaja de desarrollar cáncer de pulmón como sigue:

$$\hat{\text{ventaja}} = \frac{0,01447}{1 - 0,01447} = 0,01468 \text{ entre fumadores}$$

$$\hat{\text{ventaja}} = \frac{0,00644}{1 - 0,00644} = 0,00648 \text{ entre ex fumadores}$$

La razón de ventajas estimada es

$$\hat{\theta} = \frac{0,01468}{0,00648} = 2,265 \approx 2,3$$

Por tanto, estimamos que la ventaja de desarrollar cáncer de pulmón es aproximadamente 2,3 veces mayor para fumadores que para ex fumadores. ■

Razón de ventajas y riesgo relativo

La razón de ventajas mide la relación entre eventos de una forma poco habitual. El riesgo relativo es una medida más natural. Afortunadamente, en muchas aplicaciones las dos medidas son aproximadamente iguales. En general, la relación entre la razón de ventajas y el riesgo relativo está dada por

$$\text{razón de ventajas} = \text{riesgo relativo} \times \frac{1 - p_2}{1 - p_1}$$

Nótese que si p_1 y p_2 son pequeños, entonces el riesgo relativo es aproximadamente igual a la razón de ventajas. Ilustraremos este hecho con los datos sobre fumar y cáncer de pulmón.

Ejemplo 10.9.3 Fumar y cáncer de pulmón

Para los datos de la Tabla 10.9.1 obtuvimos que el riesgo relativo estimado de cáncer de pulmón es

$$\text{riesgo relativo estimado} = 2,247$$

y la razón de ventajas estimada es

$$\hat{\theta} = 2,265$$

Los valores son aproximadamente iguales porque el resultado de interés (contraer cáncer de pulmón) es raro, de forma que $\hat{p}_1$ y $\hat{p}_2$ son pequeños. ■

Beneficios de la razón de ventajas

Tanto el riesgo relativo p_1/p_2 como la diferencia $(p_1 - p_2)$ son más fáciles de interpretar que la razón de ventajas. ¿Por qué se utiliza entonces la razón de ventajas? Un beneficio importante de la razón de ventajas es que, en ciertos tipos de estudios, dicha razón se puede estimar incluso aunque p_1 y p_2 no se puedan estimar. Para explicar esta propiedad, debemos primero hablar de la cuestión de la posibilidad de estimar probabilidades condicionales en tablas de contingencia.

En una tabla de contingencia 2×2 , las probabilidades condicionales se pueden definir por filas o por columnas. La posibilidad de estimar estas probabilidades a partir de los datos observados depende del diseño del estudio. El ejemplo siguiente ilustra este punto.

Ejemplo 10.9.4 Fumar y cáncer de pulmón

Al estudiar la relación entre fumar y cáncer de pulmón, las probabilidades condicionales de interés principal son

$$p_1 = \Pr\{\text{cáncer de pulmón}|\text{fumador}\}$$

y

$$p_2 = \Pr\{\text{cáncer de pulmón}|\text{ex fumador}\}$$

Se trata de las probabilidades por columnas en una tabla como la Tabla 10.9.1. Sin embargo, se podrían considerar las siguientes probabilidades condicionales por filas:

$$p_1^* = \Pr\{\text{fumador}|\text{cáncer de pulmón}\}$$

y

$$p_2^* = \Pr\{\text{fumador}|\text{no cáncer de pulmón}\}$$

(Por supuesto, p_1^* y p_2^* no tienen un significado importante en sentido biológico). A partir del estudio descrito en el Ejemplo 10.9.1, es decir, una única muestra de tamaño $n = 11.900$ observada con respecto al hecho de fumar y tener cáncer de pulmón, no solo se pueden estimar p_1 y p_2 , sino también p_1^* y p_2^* . Sin embargo, hay otros diseños del estudio importantes que no proporcionan suficiente información para estimar todas estas probabilidades condicionales. Por ejemplo, supongamos que se realiza un estudio eligiendo un grupo de 500 fumadores y un grupo de 500 ex fumadores y observando cuántos de ellos contraen cáncer de pulmón. Este estudio se denomina estudio prospectivo o **estudio cohorte**. Este estudio podría producir los datos ficticios pero realistas de la Tabla 10.9.2.

Tabla 10.9.2 Datos ficticios para un estudio cohorte sobre fumar y mortalidad por cáncer de pulmón			
		Historia del fumador	
		Fumador	Ex fumador
¿Cáncer de pulmón?	Sí	7	3
	No	473	497
	Total	500	500

Los datos de la Tabla 10.9.2 se pueden ver como dos muestras independientes. A partir de los datos podemos estimar las probabilidades condicionales de contraer cáncer de pulmón en las dos poblaciones (fumadores y ex fumadores):

$$\hat{p}_1 = \frac{7}{500} = 0,014 \quad \hat{p}_2 = \frac{3}{500} = 0,006$$

Por el contrario, las probabilidades por filas de p_1^* y p_2^* no se pueden estimar a partir de la Tabla 10.9.2. Como los números relativos de fumadores y de ex fumadores están predeterminados por el diseño del estudio ($n_1 = 500$ y $n_2 = 500$), los datos no contienen información sobre la preponderancia del hecho de fumar y, por lo tanto, no hay información sobre los valores poblacionales de

$$\Pr\{\text{fumador}|\text{cáncer de pulmón}\} \quad \text{y} \quad \Pr\{\text{fumador}|\text{no cáncer de pulmón}\}$$

La Tabla 10.9.2 se ha generado fijando los totales por columnas y observando la variable de las filas. Consideremos ahora el diseño inverso. Supongamos que escogemos 500 hombres que fallecieron de cáncer de pulmón

y 500 hombres que no fallecieron de cáncer de pulmón y determinamos los historiales de fumador de dichos hombres. Este diseño se denomina **diseño de control de casos**. Dicho diseño podría generar los datos ficticios pero realistas de la Tabla 10.9.3.

Tabla 10.9.3 Datos ficticios para un estudio cohorte sobre fumar y mortalidad por cáncer de pulmón				
		Historia del fumador		Total
		Fumador	Ex fumador	
¿Cáncer de pulmón?	Sí	273	227	500
	No	173	327	500

A partir de la Tabla 10.9.3 podemos estimar las probabilidades condicionales por filas

$$\hat{p}_1^* = \frac{273}{500} = 0,546 \approx 0,55$$

$$\hat{p}_2^* = \frac{173}{500} = 0,346 \approx 0,35$$

Sin embargo, utilizando los datos de la Tabla 10.9.3 no podemos estimar las probabilidades condicionales por columnas p_1 y p_2 . Debido a que los totales por filas están predeterminados por el diseño, los datos no contienen información sobre $\Pr\{\text{cáncer de pulmón}|\text{fumador}\}$ y $\Pr\{\text{cáncer de pulmón}|\text{ex fumador}\}$. ■

El ejemplo anterior muestra que, dependiendo del diseño, un estudio puede no permitir la estimación completa de las probabilidades por columnas p_1 y p_2 , y las probabilidades por filas p_1^* y p_2^* . Afortunadamente, la razón de ventajas es la misma si se determina por filas o por columnas. Concretamente,

$$\theta = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} = \frac{\frac{p_1^*}{1-p_1^*}}{\frac{p_2^*}{1-p_2^*}}$$

Debido a esta relación, la razón de ventajas θ se puede estimar estimando a su vez p_1 y p_2 o estimando p_1^* y p_2^* . Este hecho tiene importantes aplicaciones, especialmente en estudios de control de casos, como se ilustra en el siguiente ejemplo.

Ejemplo 10.9.5 Fumadores y cáncer de pulmón

Para caracterizar la relación entre el hecho de fumar y la mortalidad por cáncer de pulmón, las probabilidades por columnas p_1 y p_2 tienen un mayor sentido biológico que las probabilidades por filas p_1^* y p_2^* . Si investigamos la relación utilizando un diseño de control de casos, ni p_1 ni p_2 se pueden estimar a partir de los datos. (Véase el Ejemplo 10.9.4). Sin embargo, la razón de ventajas *se puede* estimar a partir de los datos. Por ejemplo, utilizando la Tabla 10.9.3 obtenemos

$$\theta = \frac{\frac{\hat{p}_1^*}{1-\hat{p}_1^*}}{\frac{\hat{p}_2^*}{1-\hat{p}_2^*}} = \frac{\frac{0,546}{1-0,546}}{\frac{0,346}{1-0,346}} = 2,265 \approx 2,27$$

Podemos interpretar esta razón de ventajas como sigue: sabemos que el suceso resultado (contraer cáncer de pulmón) es raro, y sabemos también que la razón de ventajas es aproximadamente igual al riesgo relativo, p_1/p_2 .

Por tanto, estimamos que el riesgo de contraer cáncer de pulmón es aproximadamente 2,3 veces mayor en fumadores que en ex fumadores. ■

Hay una forma más fácil de calcular la razón de ventajas en una tabla de contingencia de 2×2 . En una tabla general de 2×2 , denominamos n_{11} al número de observaciones en la primera fila y en la primera columna. Asimismo, denominamos n_{12} al número de observaciones en la primera fila y en la segunda columna, y así sucesivamente. La tabla general de 2×2 tiene entonces la forma

n_{11}	n_{12}
n_{21}	n_{22}

La razón de ventajas estimadas a partir de la tabla es

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

Ejemplo 10.9.6 Fumar y cáncer de pulmón

A partir de los datos de la Tabla 10.9.1, podemos calcular la razón de ventajas como

$$\hat{\theta} = \frac{89 \times 5.711}{37 \times 6.063} = 2,265 \approx 2,27$$

El diseño de control de casos es a menudo el diseño más eficiente para investigar eventos raros, tales como enfermedades raras. Aunque la Tabla 10.9.3 se ha construido suponiendo que las dos muestras, los casos y los controles, se eligen independientemente, un diseño más común es incorporar el emparejamiento de los casos y los controles con respecto a factores de confusión potenciales (por ejemplo, la edad). Como hemos visto, aprovechando la razón de ventajas, se puede estimar el riesgo relativo de un evento raro en un estudio de control de casos incluso aunque no se puedan estimar de forma separada los riesgos p_1 y p_2 .

Si la razón de ventajas (o el riesgo relativo) es igual a 1,0, entonces las ventajas (o el riesgo) valen lo mismo en los dos grupos que se están comparando. En los datos de fumar y cáncer de pulmón de la Tabla 10.9.1, la razón de ventajas calculada eran *mayor* que 1,0, lo que indica que la ventaja del cáncer de pulmón es mayor en los fumadores que los ex fumadores. Nótese que podríamos haber enfocado nuestra atención en la ventaja de no contraer cáncer de pulmón. En este caso, la razón de ventajas sería *menor* que 1,0, como se muestra en el Ejemplo 10.9.7.

Ejemplo 10.9.7 Fumar y cáncer de pulmón

Supongamos que redispone los datos de la Tabla 10.9.1 poniendo a los que padecen cáncer de pulmón en la segunda fila y los que no lo padecen en la primera fila:

		Historia del fumador	
		Fumador	Ex fumador
¿Cáncer de pulmón?	Sí	6.063	5.711
	No	89	37
	Total	6.152	5.748

En este caso la razón de ventajas es la ventaja de no padecer cáncer de pulmón para un fumador dividida por la ventaja de no padecer cáncer de pulmón para un ex fumador. Podemos calcular la razón de ventajas estimada como

$$\hat{\theta} = \frac{6.063 \times 37}{5.711 \times 89} = 0,44$$

Este resultado es el inverso de la razón de ventajas calculada en el Ejemplo 10.9.6: $\frac{1}{2,27} = 0,44$. El hecho de que la razón de ventajas sea menor que 1,0 significa que el suceso (no padecer cáncer de pulmón) es menos probable para fumadores que para ex fumadores. ■

Intervalo de confianza para la razón de ventajas

En el Capítulo 9 presentamos el intervalo de confianza para proporciones, que es de la forma $\tilde{p} \pm z_{\alpha/2} \text{ET}_{\tilde{p}}$, siendo $\tilde{p} = \frac{y+2}{n+4}$. En concreto, el intervalo de confianza del 95 % para p se expresa como $\tilde{p} \pm z_{0,025} \text{ET}_{\tilde{p}}$. Estos intervalos de confianza se basan en el hecho de que para muestras grandes la distribución en el muestreo de $\tilde{p}$ es aproximadamente normal (de acuerdo con el Teorema Central del Límite).

De forma similar, podemos construir un intervalo de confianza para la razón de ventajas. Un problema es que la distribución muestral de $\hat{\theta}$ no es normal. Sin embargo, si tomamos el logaritmo natural de $\hat{\theta}$, entonces tenemos una distribución que es aproximadamente normal. Por tanto, podemos construir un intervalo de confianza para θ calculando primero un intervalo de confianza para $\ln(\theta)$ y transformando después los extremos para volver a la escala original.

Para poder construir un intervalo de confianza para $\ln(\theta)$ es necesario conocer el error típico de $\ln(\hat{\theta})$. La fórmula del error típico de $\ln(\hat{\theta})$ se presenta en el cuadro siguiente.

Error típico de $\ln(\hat{\theta})$

$$\text{ET}_{\ln(\hat{\theta})} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

El intervalo de confianza del 95 % para $\ln|\theta|$ se expresa como $\ln(\hat{\theta}) \pm (1,96) \text{ET}_{\ln(\hat{\theta})}$. Después se toma la exponencial de los dos extremos del intervalo, para obtener un intervalo de confianza del 95 % para θ . Los intervalos para otros niveles de confianza se construyen de forma análoga. Por ejemplo, para un intervalo de confianza del 90 % utilizaríamos $z_{0,05}$ (1,645) en vez de $z_{0,025}$ (1,960). El proceso para obtener un intervalo de confianza para θ se resume en el recuadro siguiente*.

Intervalo de confianza para θ

1. Calcular $\ln(\hat{\theta})$.
2. Construir un intervalo de confianza para $\ln(\theta)$ utilizando la fórmula $\ln(\hat{\theta}) \pm (1,96) \text{ET}_{\ln(\hat{\theta})}$.
3. Tomar la exponencial de los dos extremos del intervalo, obteniéndose así un intervalo de confianza para θ .

Este proceso se ilustra en los siguientes ejemplos.

Ejemplo 10.9.8 Fumar y cáncer de pulmón

A partir de la Tabla 10.9.1, la razón de ventajas estimada es

$$\hat{\theta} = \frac{89 \times 5.711}{37 \times 6.063} = 2,27$$

Por tanto, $\ln(\hat{\theta}) = \ln(2,27) = 0,820$.

El error típico está dado por $\text{ET}_{\ln(\hat{\theta})} = \sqrt{\frac{1}{89} + \frac{1}{37} + \frac{1}{6.063} + \frac{1}{5.711}} = 0,1965$.

El intervalo de confianza del 95 % para $\ln(\theta)$ es $0,820 \pm (1,96)(0,1965)$ o $0,820 \pm 0,385$. El intervalo es (0,435, 1,205).

* Se puede encontrar, de una forma adecuadamente modificada, un intervalo de confianza para el riesgo relativo en aquellas situaciones en las que dicho riesgo relativo se puede estimar a partir de los datos.

Para obtener un intervalo de confianza del 95 % para θ , calculamos $e^{0,435} = 1,54$ y $e^{1,205} = 3,24$. Por tanto, tenemos una confianza del 95 % en que el valor poblacional de la razón de ventajas esté entre 1,54 y 3,24. ■

Ejemplo 10.9.9 Ataques al corazón y aspirinas

Durante el Estudio sobre la Salud Médica, 11.037 médicos se asignaron aleatoriamente para tomar diariamente 325 mg de aspirina. 104 de ellos tuvieron ataques al corazón durante el estudio. Otros 11.034 médicos se asignaron aleatoriamente para tomar un placebo. 189 de ellos tuvieron ataques al corazón. Estos datos se muestran en la Tabla 10.9.4⁵³. La razón de ventajas para comparar la probabilidad de ataque al corazón con aspirina con la de ataque al corazón con placebo es

$$\hat{\theta} = \frac{189 \times 10.933}{104 \times 10.845} = 1,832$$

Por tanto, $\ln(\hat{\theta}) = \ln(1,832) = 0,605$.

El error típico es

$$ET_{\ln(\hat{\theta})} = \sqrt{\frac{1}{189} + \frac{1}{104} + \frac{1}{10.845} + \frac{1}{10.933}} = 0,123$$

El intervalo de confianza del 95 % para $\ln(\theta)$ es $0,605 \pm (1,96)(0,123)$ o $0,605 \pm 0,241$. Este intervalo es (0,364, 0,846).

Para obtener un intervalo de confianza del 95 % para θ , calculamos $e^{0,364} = 1,44$ y $e^{0,846} = 2,33$. Por tanto, tenemos un 95 % confianza en que el valor poblacional de la razón de ventajas está entre 1,44 y 2,33. Como los ataques al corazón son relativamente raros en este conjunto de datos, el riesgo relativo es casi igual a la razón de ventajas. Por tanto, podemos decir que tenemos una confianza del 95 % en que la probabilidad de un ataque al corazón es aproximadamente entre 1,44 y 2,33 veces mayor cuando se toma el placebo que cuando se toma aspirina. ■

Tabla 10.9.4 Ataques al corazón con placebo y con aspirina

	Placebo	Aspirina
Ataque al corazón	189	104
No ataque al corazón	10.845	10.933
Total	11.034	11.037

Ejercicios 10.9.1-10.9.8

10.9.1 Para cada una de las siguientes tablas, calcule (i) el riesgo relativo y (ii) la razón de ventajas.

(a)

25	23
492	614

(b)

12	8
93	84

10.9.2 Para cada una de las siguientes tablas, calcule (i) el riesgo relativo y (ii) la razón de ventajas.

(a)

14	16
322	412

(b)

15	7
338	82

En una revisión de registros médicos de perros en 27 clínicas docentes veterinarias se encontró que la displasia de cadera era más común en los Golden Retrievers que en los Border Collies. Los datos se muestran en la tabla siguiente⁵⁴. Calcule el riesgo relativo de displasia para los Golden Retrievers en comparación con los Border Collies.

		Golden Retriever	Border Collie
¿Displasia de cadera?	Sí	3.995	221
	No	42.946	5.007
	Total	46.941	5.228

10.9.3 La displasia de cadera es una anomalía del hueso de la cadera que afecta a muchos perros de razas grandes.

10.9.4 Considere los datos del Ejercicio 10.9.3.

- (a) Calcule el valor muestral de la razón de ventajas.
 (b) Construya un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas.
 (c) Interprete el intervalo de confianza del apartado (b) en el contexto de este ejercicio.

10.9.5 Como parte del Estudio Nacional de la Salud se recogieron datos sobre accidentes laborales en miles de trabajadores americanos. La tabla siguiente resume parte de esos datos⁵⁵.

		Por cuenta	
		Autónomos	ajena
¿Accidente?	Sí	210	4.391
	No	33.724	421.502
	Total	33.934	425.893

- (a) Calcule el valor muestral de la razón de ventajas.
 (b) De acuerdo con la razón de ventajas, ¿son más o menos probables los accidentes laborales en los trabajadores autónomos que en los trabajadores por cuenta ajena?
 (c) Construya un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas.
 (d) Interprete el intervalo de confianza del apartado (c) en el contexto de este ejercicio.

10.9.6 Muchos descongestivos y supresores del apetito que se compran sin receta contienen el ingrediente fenilpropanolamina. Se realizó un estudio para investigar si este compuesto estaba relacionado con hemorragias cerebrales. El estudio encontró que 6 de 702 víctimas de hemorragia cerebral habían utilizado un supresor del apetito que contenía fenilpropanolamina, en comparación con 1 de 1.376 sujetos de un grupo de control. La tabla siguiente resume los datos⁵⁶.

- (a) Calcule el valor muestral de la razón de ventajas.
 (b) Construya un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas.

- (c) Tras saber de estos datos, algunos científicos dijeron que el estudio era «no concluyente» porque el número de usuarios de supresores del apetito con fenilpropanolamina (7 en total: 6 en un grupo y 1 en el otro) era muy pequeño. ¿Qué respondería a esos científicos?

		Hemorragia	No hemorragia
¿Supresor del apetito?	Sí	6	1
	No	696	1.375
	Total	702	1.376

10.9.7 Dos tratamientos, heparina y enoxaparina, se compararon en una prueba clínica aleatorizada doble ciego en pacientes con enfermedad de la arteria coronaria. Los sujetos se pueden clasificar en función de su respuesta positiva o negativa al tratamiento. Los datos se muestran en la tabla siguiente⁵⁷.

		Heparina	Enoxaparina
Resultado	Negativo	309	266
	Positivo	1.255	1.341
	Total	1.564	1.607

- (a) Calcule el valor muestral de la razón de ventajas.
 (b) Construya un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas.
 (c) Interprete el intervalo de confianza del apartado (b) en el contexto de este ejercicio.

10.9.8 Considere los datos del Ejercicio 10.7.1. Sabiendo que hubo 139 fracturas de cadera en 1.062 pacientes que tomaron placebo y 92 fracturas de cadera en 1.065 pacientes que tomaron ácido zolendrónico, construya un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas⁴⁵.

10.10 Resumen del test chi-cuadrado

El test chi-cuadrado se aplica a menudo en tablas de contingencia. Se resume aquí.

Resumen del test chi-cuadrado para una tabla de contingencia

Hipótesis nula:

H_0 : la variable de las filas y la variable de las columnas son independientes

Cálculo de las frecuencias esperadas:

$$e_i = \frac{(\text{Total por filas}) \times (\text{Total por columnas})}{\text{Total completo}}$$

Estadístico de contraste:

$$\chi_s^2 = \sum_{\text{todas las celdas}} \frac{(o_i - e_i)^2}{e_i}$$

Distribución nula (aproximada):

distribución χ^2 con $gl = (r - 1)(k - 1)$

siendo r el número de filas y k el número de columnas en la tabla de contingencia. Esta aproximación es adecuada si $e_i \geq 5$ para todas las celdas. Si r y k son grandes, la condición de $e_i \geq 5$ es menos crítica y la aproximación χ^2 es adecuada si la frecuencia media esperada es como mínimo 5, incluso si algunas de las celdas tienen valores menores.

Las observaciones deben ser independientes entre sí. Si se toman datos pareados en una tabla 2×2 , entonces el test de McNemar es apropiado (Sección 10.8).

Ejercicios suplementarios 10.S.1-10.S.20

Nota: Los ejercicios precedidos por un asterisco se refieren a secciones opcionales).

10.S.1 En la Prueba de Modificación de la Dieta para la Iniciativa de la Salud Femenina, una serie de mujeres se asignaron aleatoriamente a un grupo de intervención y a un grupo de control. El grupo de intervención incluía sesiones de asesoramiento para reducir la ingestión de grasas y aumentar el consumo de frutas y vegetales. Se recogieron datos correspondientes a un intervalo de seis años sobre enfermedades coronarias, que se muestran en la tabla⁵⁸. ¿Proporcionan los datos evidencia de que la intervención supuso una diferencia? El valor del estadístico chi-cuadrado para esta tabla de contingencia es de $\chi^2 = 0,69$. Realice el test chi-cuadrado contra una alternativa no direccional utilizando $\alpha = 0,10$.

		Grupo		
		Intervención	Control	
¿Enfermedad coronaria?	Sí	1.000	1.549	2.549
	No	18.541	27.745	46.286
Total		19.541	29.294	48.835

10.S.2 Utilice la Tabla de Ejercicio 10.S.1 para construir un intervalo de confianza del 95 % para $\Pr\{\text{Enfermedad coronaria}|\text{Intervención}\} - \Pr\{\text{Enfermedad coronaria}|\text{Control}\}$.

10.S.3 Como parte de un estudio de las influencias del entorno en la determinación del sexo del pez *Menidia*, se tomaron huevos de un único apareamiento y se dividieron en dos grupos que se criaron en un entorno caliente o frío. Se obtuvo que 73 de los 141 descendientes del entorno caliente y 107 de los 169 descendientes del entorno frío eran hembras⁵⁹. En cada uno de los siguientes tests chi-cuadrado, utilice una alternativa no direccional y $\alpha = 0,05$.

- Contraste la hipótesis de que la proporción poblacional de sexos es 1:1 en el entorno caliente.
- Contraste de hipótesis de que la proporción poblacional de sexos es 1:1 en el entorno frío.

- Contraste de hipótesis de que la proporción poblacional de sexos es la misma en los entornos caliente y frío.
- Defina la población a la que aplican las conclusiones alcanzadas en los apartados (a)-(c). (¿Es el género *Menidia* completo?).

10.S.4 Los cilios son estructuras como cabellos que cubren la nariz y ayudan a proteger el tracto respiratorio del polvo y las partículas extrañas. Un equipo médico obtuvo especímenes de tejido nasal de niños de un jardín de infancia que tenían infecciones del tracto respiratorio superior y también de niños sanos de la misma clase. El tejido se seccionó y se examinaron los cilios con un microscopio en busca de efectos específicos, con los resultados que se muestran en la tabla que acompaña al ejercicio⁶⁰. Los datos muestran que el porcentaje de cilios defectuosos fue mucho mayor en el tejido de los niños con infección (15,7 % frente a 3,1 %). ¿Sería válido aplicar un test chi-cuadrado para comparar estos porcentajes? Si es así, hágalo. Si no, explique por qué.

		Cilios con defectos		
		Número total de cilios contabilizados	Número	Porcentaje
Control	7	556	17	3,1
Infección respiratoria	22	1.493	235	15,7

10.S.5 Un grupo de alpinistas participó en un estudio para investigar la utilidad de la sustancia acetazolamida en la prevención del mal de altura. Los alpinistas se asignaron aleatoriamente para recibir o la sustancia o el placebo durante un ascenso al Mt. Rainier. Se suponía que el experimento era doble ciego, pero surgió la cuestión de si alguno de los alpinistas tenía alguna idea (quizá a partir de la presencia o ausencia de efectos colaterales o a partir de un efecto terapéutico percibido o la ausencia del mismo) sobre qué tratamiento estaba recibiendo. Para investigar esta posibilidad, se preguntó a los alpinistas (tras terminar el experimento) que intentaran adivinar qué tratamiento habían recibido⁶¹. Los

resultados se presentan en la siguiente tabla de contingencia, para la que $\chi^2_s = 5,07$:

		Tratamiento recibido	
		Sustancia	Placebo
Adivinan	Correctamente	20	12
	Incorrectamente	11	21

Alternativamente, los mismos resultados se pueden disponer en la siguiente tabla de contingencia, para la que $\chi^2_s = 0,01$:

		Tratamiento recibido	
		Sustancia	Placebo
Adivinan	Sustancia	20	21
	Placebo	11	12

Considere la hipótesis nula

H_0 : el estudio doble ciego fue perfecto (los alpinistas no tenían idea del tratamiento).

Realice el test chi-cuadrado de H_0 contra la alternativa de que los alpinistas sí tenían idea. Utilice $\alpha = 0,05$. (Debe decidir qué tabla de contingencia es relevante para esta cuestión). (Sugerencia: para clarificar ese aspecto, intente inventar un conjunto de datos ficticio en el cual la mayoría de los alpinistas han recibido fuertes indicios, de forma que cabría esperar un valor grande de χ^2_s . Después reorganice sus datos ficticios en cada uno de los dos formatos de tabla de contingencia y vea qué tabla produciría un valor mayor de χ^2_s).

***10.S.6** Los lagartos del desierto (*Dipsosaurus dorsalis*) regulan su temperatura corporal tendiéndose al sol o moviéndose a la sombra, según lo necesitan. Normalmente, los lagartos mantienen una temperatura de aproximadamente 38 °C durante el día. Sin embargo, cuando están enfermos, mantienen una temperatura entre 2° y 4° más alta (es decir, una «fiebre»). En un experimento para ver si esta fiebre podría ser beneficiosa, se inculó a unos lagartos una infección bacteriana. Seguidamente, se evitó que 36 animales desarrollaran fiebre manteniéndolos en un recinto a 38°, mientras que otros 12 animales se mantuvieron a una temperatura de 40°. La tabla siguiente describe la mortalidad tras 24 horas⁶². ¿Con qué fuerza dan soporte estos resultados a la hipótesis de que la fiebre aumenta la supervivencia? Utilice el test exacto de Fisher contra una alternativa direccional. Use $\alpha = 0,05$.

	38°	40°
Mueren	18	2
Sobreviven	18	10
Total	36	12

10.S.7 Considere los datos del Ejercicio 10.S.6. Analice estos datos con un test chi-cuadrado. Utilice $\alpha = 0,05$.

10.S.8 En una prueba clínica aleatorizada, 154 mujeres con cáncer de pecho se asignaron para recibir quimioterapia. Otras 164 mujeres se asignaron para recibir quimioterapia combinada con terapia de radiación. Los datos de supervivencia después de 15 años se muestran en la tabla siguiente⁶³. Utilice estos datos para realizar un contraste de la hipótesis nula de que el tipo de tratamiento no afecta al porcentaje de supervivencia. Use $\alpha = 0,05$.

	Solo quimioterapia	Quimioterapia y terapia de radiación
Fallecen	78	66
Sobreviven	76	98
Total	154	164

***10.S.9** En referencia a los datos del Ejercicio 10.S.8:

- Calcule la razón de ventajas muestral.
- Calcule un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas.

10.S.10 Dos medicamentos, zidovudina y la didanosina, se probaron para ver su efectividad en la prevención del progreso del VIH en niños. En una prueba clínica doble ciego, 276 niños con VIH recibieron zidovudina, 281 recibieron didanosina y 274 recibieron zidovudina y didanosina. La tabla siguiente muestra los datos de supervivencia de los tres grupos⁶⁴. Utilice estos datos para realizar un contraste de la hipótesis nula de que la supervivencia y el tratamiento son independientes. Utilice $\alpha = 0,10$.

	Zidovudina	Didanosina	Zidovudina y didanosina
Fallecidos	17	7	10
Supervivientes	259	274	264
Total	276	281	274

10.S.11 Se compararon los grupos sanguíneos de pacientes de malaria en una clínica de la India con los obtenidos en una muestra de visitantes de un hospital cercano. Los datos se muestran en la tabla siguiente⁶⁵. Utilice estos datos para realizar un contraste de la hipótesis nula de que el grupo sanguíneo es independiente del hecho de contraer la malaria. Utilice $\alpha = 0,05$.

	A	B	O	AB	Total
Casos de malaria	138	199	106	33	476
Controles	229	535	428	96	1.300

10.S.12 El comportamiento de elección de hábitat de la mosca de la fruta *Drosophila subobscura* fue estudiado capturando moscas de dos hábitats diferentes. Las moscas se marcaron con polvo fluorescente coloreado para indicar el lugar de captura y después se liberaron en un punto intermedio entre los hábitats originales. En los dos días siguientes, se volvieron a capturar las moscas en los dos lugares. Los resultados se muestran en la tabla⁶⁶. El valor del estadístico chi-cuadrado de esta tabla de contingencia es $\chi_s^2 = 10,44$. Contraste la hipótesis nula de independencia contra la alternativa de que las moscas prefieren volver a su lugar de captura. Utilice $\alpha = 0,01$.

		Lugar de recaptura	
		I	II
Lugar de captura original	I	78	56
	II	33	58

10.S.13 En el guisante *Pisum sativum*, el color de las semillas puede ser amarillo (A) o verde (V), y la forma de la semilla puede ser redonda (R) o arrugada (U). Considere las siguientes tres hipótesis que describen una población de plantas:

$$H_0^{(1)}: \Pr\{A\} = \frac{3}{4}$$

$$H_0^{(2)}: \Pr\{R\} = \frac{3}{4}$$

$$H_0^{(3)}: \Pr\{R|A\} = \Pr\{R|V\}$$

La primera hipótesis afirma que las plantas amarillas y verdes aparecen de una proporción de 3:1. La segunda hipótesis afirma que las plantas redondas y arrugadas aparecen en una proporción de 3:1. La tercera hipótesis afirma que el color y la forma son independientes. (De hecho, en una población de plantas producida por un cierto cruce, denominado cruce dihíbrido, se sabe que las tres hipótesis son ciertas).

Suponga que se va a observar una muestra aleatoria de 1.600 plantas, con los datos dispuestos como se indica en la siguiente tabla de contingencia:

		Color	
		A	V
Forma	R		
	U		
		1.600	

Invente conjuntos de datos ficticios con las condiciones especificadas, y verifique cada respuesta calculando las probabilidades estimadas condicionales. (*Sugerencia*: en todos los casos, empiece con las frecuencias marginales).

- Un conjunto de datos que se ajusta perfectamente con $H_0^{(1)}$, $H_0^{(2)}$ y $H_0^{(3)}$.
- Un conjunto de datos que se ajusta perfectamente con $H_0^{(1)}$ y $H_0^{(2)}$, pero no con $H_0^{(3)}$.
- Un conjunto de datos que se ajusta perfectamente con $H_0^{(3)}$, pero no con $H_0^{(1)}$ ni con $H_0^{(2)}$.

***10.S.14** En un estudio de 36.080 personas que habrían sufrido ataques cardíacos se encontró que los hombres tenían más probabilidades de sobrevivir que las mujeres. La tabla siguiente muestra algunos datos recogidos en el estudio⁶⁷.

		Hombre	Mujer
¿Sobrevivieron al menos 24 horas?	Sí	25.339	8.914
	No	1.141	686
	Total	26.480	9.600

- Calcule la razón de ventajas para comparar la supervivencia de los hombres con la supervivencia de las mujeres.
- Calcule un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas.
- ¿Proporciona la razón de ventajas una buena aproximación al riesgo relativo de estos datos? ¿Por qué o por qué no?

***10.S.15** En el estudio descrito en el Ejercicio 10.9.6, una de las variables medidas fue si los sujetos habían utilizado productos conteniendo fenilpropanolamina. La razón de ventajas calculada fue de 1,49, siendo más probable que las víctimas de ataque hubieran usado un producto que contenía fenilpropanolamina que los sujetos de control⁵⁶. Un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas es (0,84, 2,64). Interprete dicho intervalo de confianza en el contexto de este ejercicio.

10.S.16 (Ejercicio para computador) En un estudio sobre los efectos de fumar cigarrillos durante el embarazo, unos investigadores examinaron la placenta de 58 mujeres después de dar a luz. Anotaron la presencia o ausencia (P o A) de una anomalía placentaria particular: vellosidades atrofiadas. Además, cada mujer se clasificó como no fumadora (N), fumadora moderada (M) o gran fumadora (G). La tabla de la página siguiente muestra, para cada mujer, un número de identificación (#), junto con los resultados de fumar (F) y vellosidades atrofiadas (V)⁶⁸.

- Contraste la relación entre el tipo de fumadora y las vellosidades atrofiadas. Utilice un test chi-cuadrado con $\alpha = 0,05$.
- Prepare una tabla que muestre el número total de mujeres en cada tipo de fumadora y el número y porcentaje en cada categoría de vellosidades atrofiadas.
- ¿Qué patrón aparece en la tabla del apartado (b) que no se haya utilizado en el test del apartado (a)?

#	F	V	#	F	V	#	F	V	#	F	V
1	N	A	16	G	P	31	M	A	46	M	A
2	M	A	17	G	P	32	M	A	47	G	P
3	N	A	18	N	A	33	N	A	48	G	P
4	M	A	19	M	P	34	N	A	49	G	A
5	M	A	20	N	P	35	N	A	50	N	P
6	M	P	21	M	A	36	G	P	51	N	A
7	G	P	22	G	A	37	N	A	52	M	P
8	N	A	23	M	P	38	G	P	53	M	A
9	N	A	24	N	A	39	G	P	54	G	P
10	M	P	25	N	P	40	N	A	55	G	A
11	N	A	26	N	A	41	M	A	56	M	P
12	N	P	27	N	A	42	N	A	57	G	P
13	G	P	28	M	P	43	G	A	58	G	P
14	M	A	29	N	A	44	M	A			
15	M	P	30	N	A	45	M	P			

***10.S.17** Unos investigadores estudiaron los registros de telefonía móvil de 699 personas que tuvieron accidentes de automóvil. Determinaron que 170 de los 699 habían realizado una llamada de teléfono móvil durante los 10 minutos previos a su accidente. Este periodo se denomina intervalo de riesgo. Hubo 37 personas que habían realizado una llamada durante el correspondiente periodo de 10 minutos en el día anterior al accidente. Este periodo se denomina intervalo de control. Finalmente hubo 13 que hicieron llamadas tanto durante el intervalo de riesgo como durante el intervalo de control⁶⁹. ¿Indican estos datos que el uso de un teléfono móvil está asociado con un incremento en el porcentaje de accidentes? Analice estos datos utilizando el test de McNemar. Utilice una alternativa direccional y $\alpha = 0,01$.

		¿Llamada durante el intervalo de control?	
		Sí	No
¿Llamada durante el intervalo de riesgo?	Sí	13	157
	No	24	505

10.S.18 Previamente a la estación de la gripe, unos sujetos se asignaron aleatoriamente para recibir o bien una vacuna de la gripe o bien un placebo. Durante esa estación hubo 28 casos de gripe entre los 813 receptores de la vacuna y 35 casos de gripe entre los 325 sujetos que recibieron el placebo⁷⁰. ¿Indican estos datos que la vacuna fue efectiva? Realice un test apropiado utilizando una alternativa direccional con $\alpha = 0,05$.

***10.S.19** En referencia a los datos del Ejercicio 10.S.18:

(a) Calcule la razón de ventajas muestral.

(b) Calcule un intervalo de confianza del 95 % para el valor poblacional de la razón de ventajas.

10.S.20 Considere el Ejercicio 9.S.18. Las parejas de los 36 hombres de la prueba de dicho ejercicio se sometieron también a dicha prueba (es decir, se les vendaron los ojos y se les pidió que identificaran a su pareja tocando las espaldas y las manos de tres hombres, uno de los cuales era su pareja). Entre las mujeres, 25 tuvieron éxito y 11 no. ¿Son estos datos evidencia significativa a favor de la hipótesis de que los hombres y las mujeres difieren en su capacidad de reconocer a su pareja?

Realice un contraste, utilizando $\alpha = 0,05$. Utilice una alternativa no direccional.

Notas

1. Guyuron, B., Reed, D., Krieglger, J., Davis, J., Pashmini, N. y Amini, S. (2009). A placebo-controlled surgical trial of the treatment of migraine headaches. *Plastic and Reconstructive Surgery* **124**, 461-468.
2. Datos no publicados cortesía de D. Wallace, recogidos en Oberlin College en el otoño de 1995.
3. Brodie, E. D., Jr. y Brodie, E. D. III. (1980). Differential avoidance of mimetic salamanders by free-ranging birds. *Science* **208**, 181-182. Copyright 1980 de la AAAS.
4. Karban, R., Adamchak, R. y Schnathorst, W. C. (1987). Induced resistance and interspecific competition between spider mites and a vascular wilt fungus. *Science* **235**, 678-680. Copyright 1987 de la AAAS.
5. Inskip, P. D., Targone, R. E., Hatch, E. E., Wilcosky, T. C., Shapiro, W. R., Selker, R. G., Fine, H. A., Black, P. M., Loeffler, J. S. y Linet, M. S. (2001). Cellular-telephone use and brain tumors. *New England Journal of Medicine* **344**, 79-86. Los datos están tomados de la Tabla 4 del artículo.
6. Turnbull, D. M., Rawlins, M. D., Weightman, D. y Chadwick, D. W. (1982). A comparison of phenytoin y valproate in previously untreated adult epileptic patients. *Journal of Neurology, Neurosurgery, and Psychiatry* **45**, 55-59.
7. Datos no publicados cortesía de W. Singleton y K. Hendrix.
8. Mizutani, T. y Mitsuoka, T. (1979). Effect of intestinal bacteria on incidence of liver tumors in gnotobiotic C3H/He male mice. *Journal of the National Cancer Institute* **63**, 1365-1370.
9. Selawry, O. S. (1974). The role of chemotherapy in the treatment of lung cancer. *Seminars in Oncology* **1**, 259-272.
10. Kannus, P., Parkkari, J., Niemi, S., Pasanen, M., Palvanen, M., Jarvinen, M. y Vuori, I. (2000). Prevention of hip fracture in elderly people with use of a hip protector. *New England Journal of Medicine* **343**, 1506-1513.
11. Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S. y Gwaltney, J. M. (1997). Social ties and susceptibility to the common cold. *Journal of the American Medical Association* **277**, 1940-1944.
12. Sherman, D. G., Atkinson, R. P., Chippendale, T., Levin, K. A., Ng, K., Futrell, N., Hsu, C. Y. and Levy, D. E. (2000). Intravenous anisodamine for treatment of acute ischemic stroke. *Journal of the American Statistical Association* **283**, 2395-2403.
13. Gueguen, N. (2009). The receptivity of women to courtship solicitation across the menstrual cycle: A field experiment. *Biological Psychology* **80**, 321-324.
14. Adaptado de Ammon, O. (1899). *Zur Anthropologie der Badener*. Jena: G. Fischer. Los datos de Ammon aparecen en Goodman, L. A. y Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association* **49**, 732-764. El cabello claro podía ser rubio o rojo; el cabello oscuro podía ser marrón o negro. Los ojos claros eran azules, grises o verdes; los ojos oscuros eran marrones.
15. Cruz-Coke, R. (1970). *Color Blindness; An Evolutionary Approach*. Springfield, Ill.: Thomas.
16. Bill-Alexson, A., et al. para el Scandinavian Prostate Cancer Study Group No. 4 (2005). Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine* **352**, 1977-1984.
17. Adaptado de Porac, C. y Coren, S. (1981). *Lateral Preferences and Human Behavior*. New York: Springer-Verlag. Las frecuencias dadas son aproximadas, deducidas de los porcentajes de las páginas 36 y 45. Las personas con preferencia neutra se contabilizaron como zurdas.
18. Upton, G. y Fingleton, B. (1985). *Spatial Data Analysis by Example: Point Pattern and Quantitative Data*, vol. 1. New York: Wiley, p. 230. Adaptado de Diggle, P. J. (1979). Statistical methods for spatial point patterns in ecology, pp. 95-150 in *Spatial and Temporal Analysis in Ecology*, R. M. Cormack y J. K. Ord (eds.). Fairland, Md.: International Cooperative Publishing House.
19. Basado en un artículo del Writing Group for Bypass Angioplasty Revascularization Investigation (BARI) Investigators (1997). Véase: five-year clinical and functional outcome comparing bypass surgery and angioplasty in patients with multivessel coronary disease. *Journal of the American Medical Association* **277**, 715-721.
20. Estos datos son ficticios, pero las proporciones de hombres y mujeres zurdos son realistas y la independencia de los gemelos está de acuerdo con datos publicados. Véase Porac, C. y Coren, S. (1981). *Lateral Preferences and Human Behavior*. New York: Springer-Verlag, p. 36; y Morgan, M. C. y Corballis, M. J. (1978). On the biological basis of human laterality: I. Evidence for a maturational left-right gradient. *The Behavioral and Brain Sciences* **2**, p. 274.
21. Ware, J. H. (1989). Investigating therapies of potentially great benefit: ECMO. *Statistical Science* **4**, 298-306. Hay controversia sobre este experimento. Un experimento anterior utilizando un esquema de aleatorización no estándar demostró que ECMO era altamente efectiva. Por tanto, algunos estadísticos cuestionan la necesidad de este segundo experimento. Para una explicación de estos aspectos véanse los artículos de las páginas 306-340 que siguen al artículo de Ware en *Statistical Science* **4**.
22. Remus, J. K. y Zahren, L. (1995). An investigation of the influenza virus at Oberlin College. Manuscrito no publicado, Oberlin College. Este estudio involucró realmente más

estudiantes que los que se indican aquí. Por simplicidad, centramos nuestra atención en los 41 estudios antes que tuvieron al menos dos catarros en el año escolar 1994-1995.

23. Hurt, R. D., Sachs, D. P. L., Glover, E. D., Offord, K. P., Johnston, J. A., Dale, L. C., Khayrallah, M. A., Schroeder, D. R., Glover, P. N., Sullivan, C. R., Croghan, I. T. y Sullivan, P. M. (1997). A comparison of sustained-release bupropion and placebo for smoking cessation. *New England Journal of Medicine* **337**, 1195-1202.

24. Datos no publicados cortesía de B. Rogers, recogidos en el Oberlin College Conservatory of Music en la primavera de 1991.

25. Souttou, B., Juhl, H., Hackenbruck, J., Rockseisen, M., Klomp, H.-J., Raulais, D., Vigny, M. y Wellstein, A. (1998). Relationship between serum concentrations of the growth factor pleiotrophin and pleiotrophin-positive tumors. *Journal of the National Cancer Institute* **90**, 1468-1473.

26. Bent, S., Kane, C., Katsuto, S., Neuhaus, J., Hudes, E. S., Goldberg, H. y Avins, A. L. (2006). Saw palmetto for benign prostatic hyperplasia. *New England Journal of Medicine* **354**, 557-566.

27. Hogarty, G. E., Kornblith, S. J., Greenwald, D., DiBarry, A. L., Cooley, S., Ulrich, R. F., Carter, M. y Flesher, S. (1997). Three-year trials of personal therapy among schizophrenic patients with or independent of family, I: Description of study and effects on relapse rates. *American Journal of Psychiatry* **154**, 1504-1513.

28. Dreitz, V. J. (2009). Parental behavior of a precocial species: Implications for juvenile survival. *Journal of Applied Ecology* **46**, 870-878.

29. Adaptado de Ammon, O. (1899). *Zur Anthropologie der Badener*. Jena: G. Fischer. Los datos de Ammon aparecen en Goodman, L. A. y Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association* **49**, 732-764.

30. Clegg, D. O., et al. (2006). Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. *New England Journal of Medicine* **354**, 795-808.

31. Inglesfield, C. y Begon, M. (1981). Open-ground individuals and population structure in *Drosophila subobscura* Collin. *Biological Journal of the Linnean Society* **15**, 259-278.

32. Aird, I., Bentall, H. H., Mehigan, J. A. y Roberts, J. A. F. (1954). The blood groups in relation to peptic ulceration and carcinoma of colon, rectum, breast, and bronchus: An association between the ABO blood groups and peptic ulceration. *British Medical Journal* **ii**, 315-321.

33. Govind, C. K. y Pearce, J. (1986). Differential reflex activity determines claw and closer muscle asymmetry in developing lobsters. *Science* **233**, 354-356. Copyright 1986 de la AAAS.

34. LeBars, P. L., Katz, M. M., Berman, N., Itil, T. M., Freedman, A. M. y Schatzberg, A. F. (1997). A placebo-controlled, double-blind, randomized trial of an extract of *Ginkgo biloba* for dementia. *Journal of the American Medical Association* **278**, 1327-1332.

35. Datos no publicados cortesía de L. Solimine.

36. Hudson, J. I., McElroy, S. L., Raymond, N. C., Crow, S., Keck, P. E., Carter, W. P., Mitchell, J. E., Strakowski, S. M., Pope, H. G., Coleman, B. S. y Jonas, J. M. (1998). Fluvoxamine in the treatment of binge-eating disorder: A multicenter placebo-controlled, double-blind trial. *American Journal of Psychiatry* **155**, 1756-1762. La variable de respuesta tenía categorías ordenadas de forma que hay métodos más potentes, fuera del alcance de este libro, que se pueden usar para analizar los datos.

37. Boden, W. E., O'Rourke, R. A., Teo, K. K., Hartigan, P. M., Maron, D. J., Kostuk, W. J., et al. COURAGE Trial Research Group. (2007). Optimal medical therapy with or without PCI for stable coronary disease. *New England Journal of Medicine* **356**, 1503-1516.

38. Wolfson, J. L. (1987). Impact of *Rhizobium* nodules on *Sitona hispidulus*, the clover root curculio. *Entomologia Experimentalis et Applicata* **43**, 237-243. Datos cortesía del autor. El experimento incluyó realmente 11 placas.

39. Adaptado de Paige, K. N. y Whitham, T. G. (1985). Individual and population shifts in flower color by scarlet gilia: A mechanism for pollinator tracking. *Science* **227**, 315-317. Los datos iniciales son ficticios, pero se han construido de forma que se ajustan a los estadísticos resumen dados por Paige y Whitham.

40. Brusco, N. K., Shields, N., Taylor, N. F. y Paratz, J. (2007). A Saturday physiotherapy service may decrease length of stay in patients undergoing rehabilitation in hospital: A randomised controlled trial. *Australian Journal of Physiotherapy* **53**, 75-81.

41. Beck, S. L. y Gavin, D. L. (1976). Susceptibility of mice to audiogenic seizures is increased by handling their dams during gestation. *Science* **193**, 427-428. Copyright 1976 de la AAAS.

42. Pittet, P. G., Acheson, K. J., Wuersch, P., Maeder, E. y Jequier, E. (1981). Effects of an oral load of partially hydrolyzed wheatflour on blood parameters and substrate utilization in man. *The American Journal of Clinical Nutrition* **34**, 2438-2445.

43. Agresti, A. y Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician* **54**, 280-288. Agresti y Caffo realizaron una serie de simulaciones que mostraron que sumar 1 a cada celda produce buenas propiedades de cobertura cuando los tamaños muestrales, n_1 y n_2 , son tan bajos como 10. Cálculos no publicados realizados por J. Witmer muestran que esas buenas propiedades también se

obtienen cuando n_1 y n_2 son tan bajos como 5, suponiendo que p_1 y p_2 no estén ambas cerca de 0 ni ambas cerca de 1, en cuyo caso los intervalos son demasiado conservadores (por ejemplo, la proporción de cobertura se acerca al 100 % para un intervalo de confianza nominal del 95 %).

44. Agresti, A. Comunicación personal.

45. Lyles, K. W., *et al.* para el HORIZON Recurrent Fracture Trial (2007). Zolendronic acid and clinical hip fractures and mortality after hip fracture. *New England Journal of Medicine* **357**, 1799-1809.

46. Saunders, M. C., Dick, J. S., Brown, I. M., McPherson, K. y Chalmers, I. (1985). The effects of hospital admission for bed rest on the duration of twin pregnancy: A randomised trial. *The Lancet* **ii**, 793-795.

47. Lader, M. y Scotto, J.-C. (1998). A multicentre double-blind comparison of hydroxyzine, buspirone and placebo in patients with generalized anxiety disorder. *Psychopharmacology* **139**, 402-406. La mejora se considera como una reducción del 50 % o superior en la puntuación del Hamilton Anxiety (HAM-A). En este estudio había un tercer grupo de tratamiento, que se ignora aquí.

48. Nesheim, S. R., Shaffer, N., Vink, P., Thea, D. M., Palumbo, P., Greenberg, B., Weedon, J. y Simmons, R. J. (1996). Lack of increased risk for perinatal human immunodeficiency virus transmission to subsequent children born to infected women. *Pediatric Infectious Disease Journal* **15**, 886-890.

49. Collaborative Group for the Study of Stroke in Young Women (1973). Oral contraception and increased risk of cerebral ischemia or thrombosis. *New England Journal of Medicine* **288**, 871-878. Reimpreso con permiso.

50. Johnson, S. K. y Johnson, R. E. (1972). Tonsillectomy history in Hodgkin's disease. *New England Journal of Medicine* **287**, 1122-1125.

51. Rillich, J., Buhl, E., Schildberger, K. y Stevenson, P. A. (2009). Female crickets are driven to fight by the male courting and calling songs. *Animal Behavior* **77**, 737-742.

52. Sidney, S., Tekawa, I. S. y Friedman, G. D. (1993). A prospective study of cigarette tar yield and lung cancer. *Cancer Causes and Control* **4**, 3-10.

53. The steering committee of the Physicians' Health Study Research Group. (1988). Preliminary report: Findings from the aspirin component of the ongoing physicians' health study. *New England Journal of Medicine* **318**, 262-264.

54. Witsberger, T. H., Villamil, J. A., Schultz, L. G., Hahn, A. W. y Cook, J. L. (2008). Prevalence of and risk factors for hip dysplasia and cranial cruciate ligament deficiency in dogs. *Journal of the American Veterinary Medical Association* **232**, 1818-1824.

55. Zwerling, C., Whitten, P. S., Davis, C. S. y Sprince, N. L. (1997). Occupational injuries among workers with disabilities. *Journal of the American Medical Association*

278, 2163-2166. En este estudio daño significa daño ocupacional en el año anterior a la entrevista de la persona.

56. Kernan, W. N., Viscoli, C. M., Brass, L. M., Broderick, J. P., Brott, T., Feldmann, E., Morgenstern, L. B., Wilterdink, J. L. y Horwitz, R. I. (2000). Phenylephrine and the risk of hemorrhagic stroke. *New England Journal of Medicine* **343**, 1826-1832.

57. Cohen, M., Demers, C., Gurfinkel, E. P., Turpie, A. G., Fromell, G. J., Goodman, S., Langer, A., Califf, R. M., Fox, K. A. A., Premmureur, J. y Bigonzi, F. (1997). A comparison of low-molecular-weight heparin with unfractionated heparin for unstable coronary artery disease. *New England Journal of Medicine* **337**, 447-452. Un resultado negativo aquí es fallecimiento, infarto de miocardio o angina recurrente durante los primeros 14 días después del tratamiento.

58. Howard, B. V., *et al.* (2006). Low-fat dietary pattern and risk of cardiovascular disease. *Journal of the American Medical Association* **295**, 655-666.

59. Conover, D. O. y Kynard, B. E. (1981). Environmental sex determination: Interaction of temperature and genotype in a fish. *Science* **213**, 577-579. Copyright 1981 de la AAAS.

60. Carson, J. L., Collier, A. M. y Hu, S. S. (1985). Acquired ciliary defects in nasal epithelium of children with acute viral upper respiratory infections. *New England Journal of Medicine* **312**, 463-468. Reimpreso con permiso.

61. Larson, E. B., Roach, R. C., Schoene, R. B. y Hombein, T. F. (1982). Acute mountain sickness and acetazolamide. *Journal of the American Medical Association* **248**, 328-332. Copyright 1982 American Medical Association.

62. Kluger, M. J., Ringler, D. H. y Anver, M. R. (1975). Fever and survival. *Science* **188**, 166-168. Copyright 1975 de la AAAS. El artículo original contenía una errata, pero el Dr. Kluger proporcionó amablemente la mortalidad correcta a 40 °C.

63. Ragaz, J., Jackson, S. M., Le, N., Plenderleith, I. H., Spinelli, J. J., Basco, V. E., Wilson, K. S., Knowling, M. A., Coppin, C. M. L., Paradis, M., Coldman, A. J. y Olivotto, I. A. (1997). Adjuvant radiotherapy and chemotherapy in node-positive premenopausal women with breast cancer. *New England Journal of Medicine* **337**, 956-962.

64. Englund, J. A., Baker, C. J., Raskino, C., McKinney, R. E., Petrie, B., Fowler, M. G., Pearson, D., Gershon, A., McSherry, G. D., Abrams, E. J., Schliozberg, J. y Sullivan, J. L. (1997). Zidovudine, didanosine, or both as the initial treatment for symptomatic HIV-infected children. *New England Journal of Medicine* **336**, 1704-1712. Los datos presentados aquí son para un análisis provisional que se realizó aproximadamente dos años en el estudio. Como resultado del análisis provisional de porcentajes de fallecimientos y de progresión de la enfermedad, se interrumpió el uso de zidovudina antes del final de la prueba.

65. Gupta, M. y Chordhuri, A. N. R. (1980). Relationship between ABO blood groups and malaria. *Bulletin of the World Health Organization* **58**, 913-915.
66. Shorrocks, B. y Nigro, L. (1981). Microdistribution and habitat selection in *Drosophila subobscura* collin. *Biologic al Journal of the Linnean Society* **16**, 293-301.
67. Malacrida, R., Genoni, M., Maggioni, A. P., Spatato, V., Parish, S., Palmer, A., Collins, R. y Moccetti, T. (1998). A comparison of the early outcome of acute myocardial infarction in women and men. *New England Journal of Medicine* **338**, 8-14. Aunque la razón de ventajas para estos datos muestra que es más probable la supervivencia de los hombres que de las mujeres, los autores comentan el efecto de la edad en este hecho. Calcularon una nueva razón de ventajas tras ajustar la edad y otros factores (utilizando métodos fuera del alcance de este libro) y concluyeron que buena parte de la diferencia en la probabilidad de supervivencia se debía a esos factores.
68. Mochizuki, M., Marno, T., Masuko, K. y Ohtsu, T. (1984). Effects of smoking on fetoplacental-maternal system during pregnancy. *American Journal of Obstetrics and Gynecology* **149**, 413-420.
69. Redelmeier, D. A. y Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine* **336**, 453-458. Véase también Redelmeier, D. A. y Tibshirani, R. J. (1997). Is using a car phone like driving drunk? *Chance* **10**, n.º 2, 5-9.
70. Monto, A. S., Ohmit, S. E., Petrie, J. G., Johnson, E., Truscon, R., Teich, E., Rotthoff, J., Boulton, M. y Victor, J. (2009). Comparative efficacy of inactivated and live attenuated influenza vaccines. *New England Journal of Medicine* **361**, 1260-1267.

COMPARACIÓN DE MEDIAS DE VARIAS MUESTRAS INDEPENDIENTES

Objetivos

En este capítulo estudiaremos el análisis de varianza (ANOVA):

- explicaremos cuándo y por qué realizar un análisis de varianza;
- desarrollaremos intuitivamente las bases del modelo ANOVA;
- demostraremos cómo se realizan los cálculos del ANOVA;
- describiremos y examinaremos las condiciones bajo las que es válido un modelo ANOVA;
- veremos cómo utilizar bloques y cómo realizar un ANOVA con bloques aleatorizados;
- describiremos las interacciones y los efectos principales en modelos ANOVA factoriales;
- construiremos contrastes y otras combinaciones lineales de medias;
- introduciremos y compararemos varios métodos para tratar las comparaciones múltiples.

11.1 Introducción

En el Capítulo 7 consideramos la comparación de dos muestras independientes con respecto a la variable cuantitativa Y . Las técnicas clásicas para comparar las dos medias muestrales $\bar{Y}_1$ e $\bar{Y}_2$ son el contraste de hipótesis y el intervalo de confianza basados en la distribución t de Student. En el presente capítulo consideraremos la comparación de las medias de I muestras independientes, donde I puede ser mayor que 2. El ejemplo siguiente ilustra un experimento con $I = 5$.

Ejemplo 11.1.1 Maíz dulce

Al cultivar maíz dulce, ¿pueden utilizarse con éxito métodos orgánicos para controlar los insectos dañinos y limitar su efecto en el maíz? En un estudio de esta cuestión unos investigadores compararon los pesos de mazorcas de maíz bajo cinco condiciones en un experimento en el que se cultivó maíz dulce utilizando métodos orgánicos. En una parcela de maíz se introdujo un nematodo del suelo beneficioso. En una segunda parcela se utilizó una avispa parásita. Una tercera parcela se trató con el nematodo y la avispa. En una cuarta parcela se utilizó una bacteria. Finalmente, una quinta parcela se utilizó como control: en esta última no se aplicó ningún tratamiento. Por tanto, los tratamientos fueron:

- Tratamiento 1: nematodos.
- Tratamiento 2: avispas.
- Tratamiento 3: nematodos y avispas.
- Tratamiento 4: bacteria
- Tratamiento 5: control

Las mazorcas de maíz se muestrearon aleatoriamente en cada parcela y se pesaron. Los resultados se muestran en la Tabla 11.1.1 y se representan gráficamente en la Figura 11.1.1¹. Nótese que además de las diferencias entre las medias de los tratamientos, hay también una variación considerable dentro de cada grupo de tratamiento. ■

Tabla 11.1.1 Pesos (en onzas) de mazorcas de maíz dulce					
	Tratamiento				
	1	2	3	4	5
	16,5	11,0	8,5	16,0	13,0
	15,0	15,0	13,0	14,5	10,5
	11,5	9,0	12,0	15,0	11,0
	12,0	9,0	10,0	9,0	10,0
	12,5	11,5	12,5	10,5	14,0
	9,0	11,0	8,5	14,0	12,0
	16,0	9,0	9,5	12,5	11,0
	6,5	10,0	7,0	9,0	9,5
	8,0	9,0	10,5	9,0	18,5
	14,5	8,0	10,5	9,0	17,0
	7,0	8,0	13,0	6,5	10,0
	10,5	5,0	9,0	8,5	11,0
Media	11,5	9,6	10,3	11,1	12,3
DT	3,5	2,4	2,0	3,1	2,9
<i>n</i>	12	12	12	12	12

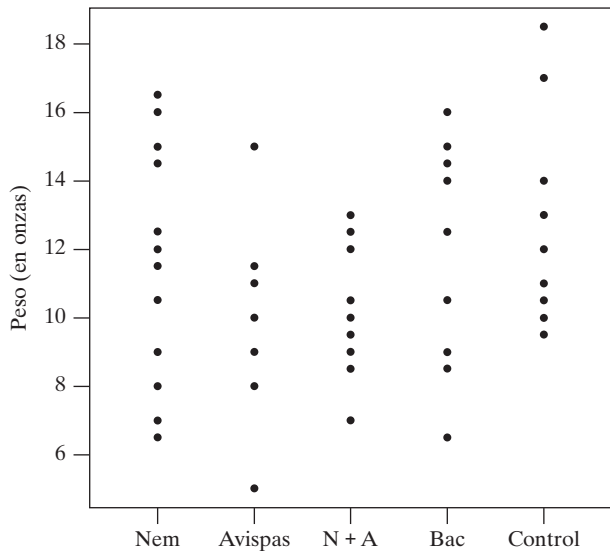


Figura 11.1.1 Pesos de mazorcas de maíz que han recibido cinco tratamientos diferentes

Comentaremos el método clásico para analizar datos de I muestras independientes. El método se denomina **análisis de varianza** o **ANOVA**. Para aplicar el análisis de la varianza, los datos se consideran una muestra aleatoria de I poblaciones. Denotaremos las medias poblacionales como $\mu_1, \mu_2, \dots, \mu_I$ y las desviaciones típicas poblacionales como $\sigma_1, \sigma_2, \dots, \sigma_I$.

¿Por qué no repetir el tests t ?

Es natural preguntarse por qué la comparación de las medias de I muestras requiere un nuevo método. Por ejemplo, ¿por qué no aplicar un test t de dos muestras a cada pareja de muestras? He aquí tres razones de por qué eso no es una buena idea.

1. *El problema de las comparaciones múltiples.* La dificultad más seria con un procedimiento simplista de «repetir tests t » tiene que ver con el error de Tipo I. La probabilidad de falso rechazo de una hipótesis nula puede ser mucho mayor de lo que parece. Por ejemplo, supongamos que $I = 4$ y consideremos la hipótesis nula de que las medias de las cuatro poblaciones son iguales ($H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$) frente a la hipótesis alternativa de que las cuatro medias no son iguales*. Hay seis posibles formas de comparar parejas de cuatro medias. Los emparejamientos se muestran en la Figura 11.1.2. Las seis hipótesis resultantes son

$$\begin{array}{lll}
 H_0: \mu_1 = \mu_2 & H_0: \mu_1 = \mu_3 & H_0: \mu_1 = \mu_4 \\
 H_0: \mu_2 = \mu_3 & H_0: \mu_2 = \mu_4 & H_0: \mu_3 = \mu_4
 \end{array}$$

Consideremos el riesgo de cometer un error de Tipo I al contrastar realizando seis tests t distintos procedentes de nuestra hipótesis nula primaria que establece que las cuatro medias son iguales. Si en cualquiera de los seis tests t obtiene una diferencia significativa entre una pareja de medias, rechazaríamos nuestra hipótesis nula primaria de

* En la Sección 11.2 elaboraremos más la forma de la hipótesis alternativa.

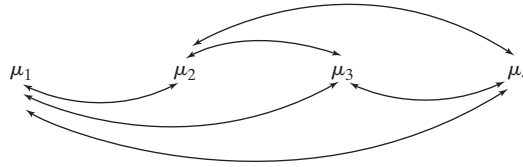


Figura 11.1.2 La comparación de las medias de cuatro poblaciones requiere realizar seis comparaciones

que las cuatro medias son iguales. Ocurriría un error de Tipo I si *cualquiera* de los seis tests obtuviera una diferencia significativa entre una pareja de medias cuando de hecho las cuatro medias son iguales. Por tanto, utilizando $\alpha = 0,05$ en cada uno de los tests t individuales se produce un riesgo total de error de Tipo I que es mayor que el 5 %.

Nuestra intuición podría sugerir que el riesgo global de un error de Tipo I en el ejemplo anterior sería $6 \times 0,05 = 0,3 = 30\%$ (en cada uno de los seis tests tenemos un 5 % de probabilidad de encontrar de forma errónea evidencia a favor de una diferencia), pero este no es el caso. El cálculo de este error de Tipo I global es más complejo. La Tabla 11.1.2 muestra el riesgo global de error de Tipo I*, que es:

Riesgo global de error de Tipo I = Probabilidad de que al menos uno de los tests t rechazara la hipótesis nula, cuando de hecho $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_I$.

Tabla 11.1.2 Riesgo global de error Tipo I al usar tests t repetidos con $\alpha = 0,05$	
I	Riesgo global
2	0,05
3	0,12
4	0,20
6	0,37
8	0,51
10	0,63

Si $I = 2$, entonces el riesgo total es 0,05, como debería ser, pero con valores de I mayores el riesgo aumenta rápidamente. Para $I = 6$ es 0,37. Observando la Tabla 11.1.2 resulta claro que el investigador que utilice tests t repetidos es altamente vulnerable al error de Tipo I a menos que I sea bastante pequeño.

Las dificultades ilustradas en la Tabla 11.1.2 son debidas a las **comparaciones múltiples**, es decir, muchas comparaciones sobre el mismo conjunto de datos. Estas dificultades se pueden reducir cuando la comparación de varios grupos se plantea mediante el ANOVA.

2. *Estimación de la desviación típica.* La técnica del ANOVA combina información sobre la variabilidad de todas las muestras simultáneamente. Al compartir la información globalmente, se puede mejorar la precisión del análisis.
3. *Estructura en los grupos.* En muchos estudios la estructura lógica de los tratamientos o los grupos que se van a comparar puede inspirar cuestiones que no se pueden responder por simples comparaciones entre parejas. Por ejemplo, podríamos desear estudiar los efectos de dos factores experimentales simultáneamente. El ANOVA se puede utilizar para analizar datos en estos casos (véanse las Secciones 11.6, 11.7 y 11.8).

* La Tabla 11.1.2 se ha calculado suponiendo que los tamaños muestrales son grandes e iguales y que las distribuciones poblacionales son normales con desviaciones típicas iguales.

Una perspectiva gráfica del ANOVA

Cuando se analizan datos mediante análisis de varianza, el primer paso usual es contrastar la siguiente hipótesis nula global:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_I$$

que asevera que todas las medias poblacionales son iguales. Un contraste estadístico de H_0 se describirá en la Sección 11.4. Sin embargo, consideraremos primero el análisis de la varianza desde una perspectiva gráfica.

Consideremos el diagrams de puntos que se muestran en la Figura 11.1.3(a). Este diagrama de puntos ha sido generado en un contexto en el cual H_0 es cierta. Las medias muestrales, que se muestran como líneas en la gráfica, difieren entre sí solo como resultado del error por azar. En los datos que se muestran en la Figura 11.1.3(b), H_0 es falsa. Las medias muestrales son bastante diferentes [hay una variabilidad sustancial entre las medias de los grupos, que proporciona evidencia de que las correspondientes medias poblacionales (μ_1 , μ_2 , μ_3 y μ_4) no son todas iguales]. En este caso particular, parece que μ_1 y μ_2 son diferentes de μ_3 y μ_4 .

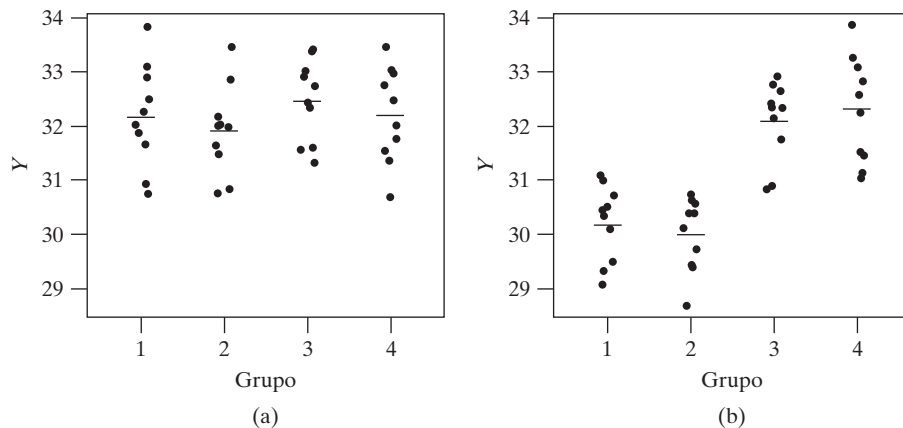


Figura 11.1.3 (a) H_0 cierta, (b) H_0 falsa, con DT pequeña en los grupos

La Figura 11.1.4 muestra una situación que es menos clara. De hecho, H_0 es falsa en este caso [las medias en la Figura 11.1.4 son idénticas a las de la Figura 11.1.3(b)]. Sin embargo, las desviaciones típicas de los grupos son bastante grandes, lo que hace difícil decidir que las medias poblacionales son diferentes*.

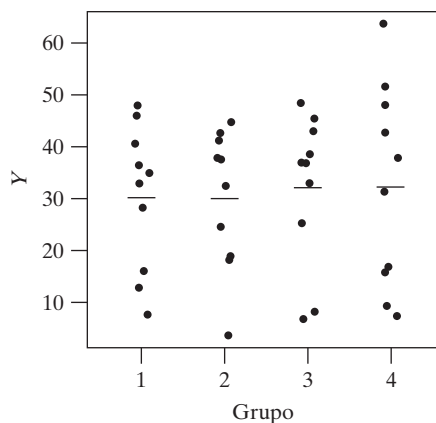


Figura 11.1.4 H_0 cierta, con DT grande en los grupos

* Nótese el cambio de escala del eje vertical en la Figura 11.1.4.

Necesitamos saber cuánta variabilidad inherente hay en los datos antes de poder juzgar si una diferencia en las medias muestrales es bastante pequeña y atribuible al azar o si es demasiado grande para ser debida solo al azar. Como ilustran las Figuras 11.1.3 y 11.1.4, para encontrar evidencia convincente a favor de una diferencia en las medias poblacionales, no solamente debe existir (1) variación entre las medias de los grupos, sino que además debe ser relativamente grande *comparada* con (2) la variabilidad inherente en los grupos. Solo mediante la comparación de las magnitudes relativas de esos dos tipos de *variabilidad* («análisis de varianza») seremos capaces de realizar una inferencia sobre las medias.

Una mirada hacia delante

Si la hipótesis nula global de $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_I$ se rechaza, entonces los datos proporcionan evidencia suficiente para concluir que al menos *alguna* de las μ no son iguales. El investigador puede proceder entonces a realizar comparaciones detalladas para determinar el *patrón* de las diferencias entre las μ . Si hay una falta de evidencia contra la hipótesis nula global, entonces el investigador puede construir uno o más intervalos de confianza para caracterizar la pérdida de diferencias significativas entre las μ .

Todos los procedimientos estadísticos de este capítulo (el test de la hipótesis nula global y varios métodos para realizar comparaciones detalladas entre las medias) dependen de los mismos cálculos básicos. Estos cálculos se presentan en la Sección 11.2.

11.2 El análisis de varianza con un factor

El modelo de ANOVA presentado en la Sección 11.1 que compara las medias de tres o más grupos se denomina **ANOVA de un factor**. El término «un factor» se refiere al hecho de que solo hay una variable que define los grupos o tratamientos (es decir, en el ejemplo del maíz dulce los tratamientos se basaban en el tipo de bacteria/insecto dañino). Posteriormente en este capítulo examinaremos otros modelos de ANOVA tales como el ANOVA por bloques completos aleatorizados (Sección 11.6) y el modelo de ANOVA de dos factores (Sección 11.7), que consideran el impacto de tener más de una variable que define los grupos o cómo se asignan los tratamientos a las unidades experimentales.

En esta sección presentamos los cálculos básicos del ANOVA de un factor que se utilizan para describir los datos y facilitar análisis posteriores. En la sección anterior comentamos que si la variabilidad de las medias entre los grupos es grande en relación con la variabilidad dentro de los grupos, tomaremos esto como evidencia contra la hipótesis nula de que las medias poblacionales son todas iguales. Por tanto, el análisis de varianza de I muestras, o grupos, comienza con el cálculo de valores que describen la variabilidad de los datos *entre* los grupos y *dentro* de los grupos o intra grupos. (Por claridad, en este capítulo nos referiremos a menudo a las muestras como «grupos» de observaciones).

Notación

Para describir varios grupos de observaciones cuantitativas, utilizaremos dos subíndices: uno para llevar la cuenta de la pertenencia al grupo y otro para llevar la cuenta de las observaciones dentro de los grupos. Por tanto, denotaremos la observación j en el grupo i como

$$y_{ij} = \text{observación } j \text{ del grupo } i$$

Por tanto, la primera observación del primer grupo es y_{11} , la segunda observación del primer grupo es y_{12} , la tercera observación del segundo grupo es y_{23} , y así sucesivamente.

Utilizaremos también la siguiente notación:

I = número de grupos

n_i = número de observaciones en el grupo i

$\bar{y}_i$ = media del grupo i

s_i = desviación típica del grupo i

El número total de observaciones es

$$n = \sum_{i=1}^I n_i$$

Finalmente, la **media global** (la media de todas las observaciones) $\bar{Y}_{\text{conjunta}}$ o simplemente $\bar{Y}$ es:

$$\bar{Y}_{\text{conjunta}} = \bar{y} = \frac{\sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}}{n}$$

Podemos expresar $\bar{y}$ de forma equivalente como una media ponderada de las medias de los grupos

$$\bar{y} = \frac{\sum_{i=1}^I n_i \bar{y}_i}{\sum_{i=1}^I n_i} = \frac{\sum_{i=1}^I n_i \bar{y}_i}{n}$$

El siguiente ejemplo ilustra esta notación.

Ejemplo 11.2.1 Ganancia de peso de corderos

La Tabla 11.2.1 muestra las ganancias de peso (en dos semanas) de corderos jóvenes bajo tres dietas diferentes. (Estos datos son ficticios, pero realistas en todos los aspectos excepto por el hecho de que las medias de los grupos son números enteros)².

El número total de observaciones es

$$n = 3 + 5 + 4 = 12$$

Tabla 11.2.1 Ganancia de peso de corderos (lb)*			
	Dieta 1	Dieta 2	Dieta 3
	8	9	15
	16	16	10
	9	21	17
		11	6
		18	
n_i	3	5	4
Suma = $\sum_{j=1}^{n_i} y_{ij}$	33	75	48
Media = $\bar{y}_i$	11,000	15,000	12,000
DT = s_i	4,359	4,950	4,967
* Los dígitos extra se utilizan para obtener precisiones de los cálculos subsiguientes.			

y el total de todas las observaciones es

$$\sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij} = 33 + 75 + 48 = 156 \text{ o, de forma equivalente } 3 \times 11 + 5 \times 15 + 4 \times 12 = 156$$

La media global es

$$\bar{y} = \frac{156}{12} = 13 \text{ lb}$$



Si los tamaños muestrales (n_i) son todos iguales, entonces la media global $\bar{y}$ es simplemente el promedio ordinario (es decir, la media) de las medias de los grupos (los $\bar{y}_i$). Pero si los tamaños muestrales no son iguales, entonces no es el caso. Por ejemplo, en el Ejemplo 11.2.1 nótese que

$$\frac{11 + 15 + 12}{3} \neq 13$$

Medida de la variación dentro de los grupos

Una medida combinada de la variación dentro de los I grupos es la desviación típica conjunta s_{conjunta} , a menudo se denomina simplemente s , que se calcula como sigue*.

Desviación típica conjunta

$$s_{\text{conjunta}} = s = \sqrt{\frac{\sum_{i=1}^I (n_i - 1)s_i^2}{\sum_{i=1}^I (n_i - 1)}} = \sqrt{\frac{\sum_{i=1}^I (n_i - 1)s_i^2}{n - I}}$$

Denominamos $s_{\text{conjunta}}^2 = s^2$ a la varianza conjunta†

$$s_{\text{conjunta}}^2 = s^2 = \frac{\sum_{i=1}^I (n_i - 1)s_i^2}{\sum_{i=1}^I (n_i - 1)}$$

Examinando la fórmula podemos ver que la varianza conjunta es una media ponderada de las varianzas muestrales de los grupos y, por tanto, la desviación típica conjunta puede ser interpretada de forma muy flexible como la media ponderada de las desviaciones típicas de los grupos.

Ejemplo 11.2.2 Ganancia de peso de corderos

La Tabla 11.2.1 muestra los tamaños muestrales y las desviaciones típicas de los grupos para los datos de ganancia de peso de corderos. La varianza y desviación típica conjunta se calculan como

$$s^2 = \frac{(3 - 1)4,359^2 + (5 - 1)4,950^2 + (4 - 1)4,967^2}{12 - 3} = \frac{210,025}{9} = 23,336$$

$$s = \sqrt{23,336} = 4,831$$

Obsérvese que la desviación típica conjunta, 4,831 lb, es un valor representativo sensato de las tres desviaciones típicas de los grupos, 4,359, 4,950 y 4,967 lb. Si suponemos que la desviación típica de la población de las ganancias de peso es la misma en las tres dietas, entonces estimaríamos este valor común como 4,83 libras. Esta estimación depende solo de la variabilidad dentro de los grupos y no de sus valores medios. La Figura 11.2.1(a) muestra los datos de la Tabla 11.2.1 mientras que la Figura 11.2.1(b) muestra una versión modificada de los datos en los que se ha añadido el valor 7 a cada una de las observaciones de la Dieta 2 y se ha sustraído el valor 5 a cada una de las observaciones de la Dieta 3. Vemos que aunque las medias de los grupos son diferentes para estos dos conjuntos de datos, la desviación típica conjunta (la variabilidad inherente en cada grupo) es la misma.

Notación del ANOVA

Aunque nuestras fórmulas anteriores utilizan notación y términos familiares, es conveniente descomponer la varianza conjunta en partes y definir subsecuentemente nuevos términos que se utilizarán en el contexto del análisis de la varianza.

* No existe ambigüedad en esta notación, ya que s_i (es decir, s con subíndice) indica la desviación típica muestral de un grupo individual.

† Recuérdese del Capítulo 2 que la varianza es simplemente la desviación típica al cuadrado.

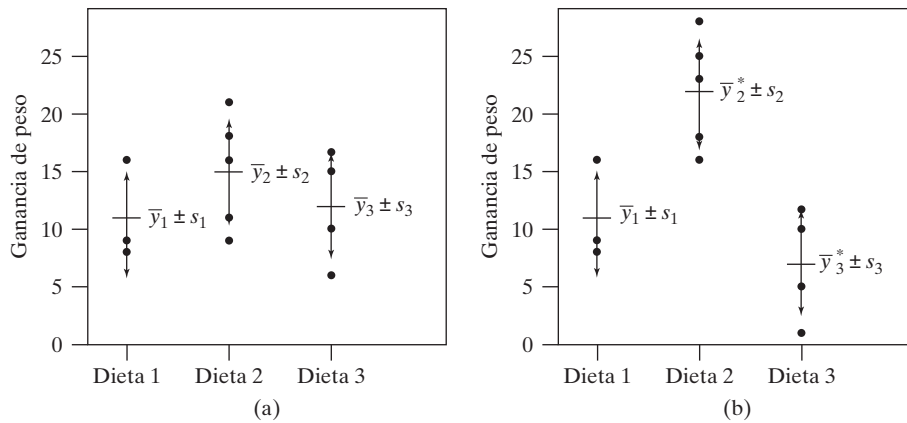


Figura 11.2.1 Examen de las desviaciones típicas dentro de los grupos. La gráfica (a) muestra los datos de ganancia de peso de la Tabla 11.2.1, con $s = 4,831$. La gráfica (b) muestra datos modificados con las mismas desviaciones típicas de los grupos individuales y, por tanto, la misma desviación típica conjunta $s = 4,831$

El numerador de la varianza conjunta se conoce como **suma de cuadrados dentro de los grupos, SC (intra)**, mientras que el denominador se conoce como **grados de libertad dentro de los grupos, gl (intra)**. Sus fórmulas se presentan en el recuadro siguiente*.

Suma de cuadrados y gl dentro de los grupos

$$SC(\text{intra}) = \sum_{i=1}^I (n_i - 1)s_i^2$$

$$gl(\text{intra}) = n - I$$

Su relación se define como **cuadrados medios dentro de los grupos o CM (intra)**. Nótese que CM (intra) es solo otro nombre de la varianza conjunta.

Cuadrados medios dentro de los grupos

$$CM(\text{intra}) = \frac{SC(\text{intra})}{gl(\text{intra})}$$

Por tanto, el valor CM(intra) mide la variabilidad dentro de los grupos[†].

El siguiente ejemplo ilustra el cálculo de SC(intra), gl(intra) y CM(intra).

Ejemplo 11.2.3 Ganancia de peso de corderos

En el Ejemplo 11.2.2, cuando calculamos la varianza conjunta, obtuvimos

$$s^2 = \frac{(3 - 1)4,359^2 + (5 - 1)4,950^2 + (4 - 1)4,967^2}{12 - 3} = \frac{210,025}{9} = 23,336$$

Por tanto, $SC(\text{intra}) = 210,025$, $gl(\text{intra}) = 9$ y $CM(\text{intra}) = 23,336$. ■

* Una fórmula popular pero menos intuitiva de SC(intra) es $SC(\text{intra}) = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$.

† Si solo hubiera un grupo, con n observaciones, entonces gl(intra) sería $n - 1$ y la SC(intra) sería $(n - 1)s^2$. CM(intra) sería entonces simplemente $\frac{(n - 1)s^2}{(n - 1)} = s^2$, la varianza muestral.

Variación entre grupos

En el caso de dos grupos, la diferencia entre los mismos se describe simplemente mediante $(\bar{y}_1 - \bar{y}_2)$. ¿Cómo podemos describir la variabilidad entre grupos para más de dos grupos? Una idea simple es calcular sencillamente la varianza muestral de las medias de los grupos. Los **cuadrados medios entre grupos, o CM(entre)** está motivado por esta idea. De hecho, si no fuera por el n_i en el numerador de la expresión siguiente (para ajustar a los tamaños muestrales de los grupos), el CM(entre) sería de hecho la varianza muestral de las medias de los grupos.

Cuadrados medios entre grupos

$$\text{CM(entre)} = \frac{\sum_{i=1}^I n_i (\bar{y}_i - \bar{y})^2}{I - 1}$$

Como en el caso de las medidas utilizadas para la variación dentro de los grupos, CM (intra), es conveniente definir el numerador de CM(entre) como la **suma de cuadrados entre grupos o SC(entre)** y el denominador como los **grados de libertad entre grupos o gl(entre)** de forma que

$$\text{CM(entre)} = \frac{\text{SC(entre)}}{\text{gl(entre)}}$$

donde SC(entre) y gl(entre) se definen explícitamente como sigue.

Suma de cuadrados y gl entre grupos

$$\begin{aligned} \text{SC(entre)} &= \sum_{i=1}^I n_i (\bar{y}_i - \bar{y})^2 \\ \text{gl(entre)} &= I - 1 \end{aligned}$$

El siguiente ejemplo ilustra estas definiciones.

Ejemplo 11.2.4 Ganancia de peso de corderos

Para los datos del Ejemplo 11.2.1, las cantidades que intervienen en SC(entre) se muestran en la Tabla 11.2.2.

Tabla 11.2.2 Cálculo de SC(entre) para las ganancias de peso de corderos			
	Dieta 1	Dieta 2	Dieta 3
Media: $\bar{y}_i$	11	15	12
n_i	3	5	4
Media global $\bar{y} = 13$			

De la Tabla 11.2.2 calculamos

$$\text{SC(entre)} = 3(11 - 13)^2 + 5(15 - 13)^2 + 4(12 - 13)^2 = 36$$

Como $I = 3$, tenemos

$$\text{gl(entre)} = 3 - 1 = 2$$

de forma que

$$\text{CM(entre)} = \frac{36}{2} = 18$$



La SC (entre) y el CM (entre) mide en la variabilidad entre las medias muestrales de los grupos. Esta variabilidad se muestra gráficamente en la Figura 11.2.2.

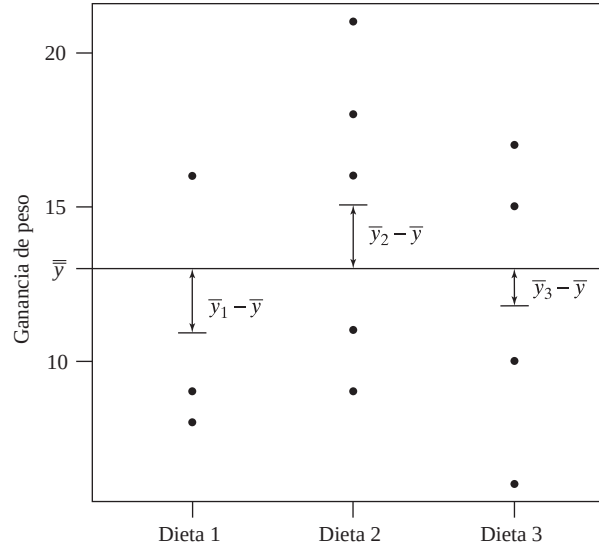


Figura 11.2.2 Medida de las diferencias entre las medias de los grupos

Una relación fundamental del ANOVA

El nombre de *análisis de la varianza* se deriva de una relación fundamental entre SC(entre) y SC(intra). Consideremos una única observación y_{ij} . Es obviamente cierto que

$$y_{ij} - \bar{y} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y})$$

Esta ecuación expresa la desviación de una observación de la media global como la suma de dos partes: una desviación dentro de los grupos ($y_{ij} - \bar{y}_i$) y una desviación entre grupos ($\bar{y}_i - \bar{y}$). Es también cierto (pero en absoluto obvio) que la relación análoga se mantiene también para las sumas de cuadrados correspondientes, es decir,

$$\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2 \quad (11.2.1)$$

que, reescribiendo las sumas del miembro derecho de la ecuación se puede expresar como

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 &= \sum_{i=1}^I (n_i - 1)s_i^2 + \sum_{i=1}^I n_i (\bar{y}_i - \bar{y})^2 \\ &= \text{SC(intra)} + \text{SC(entre)} \end{aligned}$$

La cantidad del miembro izquierdo de la fórmula (11.2.1) se denomina **suma total de cuadrados**, o **SC(total)**:

Definición de suma total de cuadrados

$$\text{SC(total)} = \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$$

Nótese que SC(total) mide la variabilidad entre todas las n observaciones de los I grupos. La relación [fórmula (11.2.1)] se puede expresar como

Relación entre las sumas de cuadrados

$$SC(\text{total}) = SC(\text{entre}) + SC(\text{intra})$$

La precedente relación fundamental muestra cómo se puede analizar, o descomponer, la variación total del conjunto de datos en dos componentes interpretables: la variación entre muestras y la variación dentro de las muestras. Esta partición es un análisis de varianza.

Los **grados de libertad totales**, o **gl(total)**, se definen como sigue:

Gl total

$$gl(\text{total}) = n - 1$$

Con esta definición, los grados de libertad se suman, como lo hace la suma de cuadrados, es decir

$$gl(\text{total}) = gl(\text{intra}) + gl(\text{entre})$$

$$n - 1 = (n - I) + (I - 1)$$

Nótese que, si consideráramos todas las observaciones n como una única muestra, entonces la SC para esa muestra (es decir, el numerador de la varianza), sería $SC(\text{total})$ y los gl asociados (es decir, el denominador de la varianza) sería $gl(\text{total})$. En consecuencia, $\sqrt{\frac{SC(\text{total})}{gl(\text{total})}}$ es la desviación típica del conjunto de datos completo cuando se ignora la pertenencia a los grupos.

El ejemplo siguiente ilustra las relaciones fundamentales entre la suma de cuadrados y los grados de libertad.

Ejemplo 11.2.5 Ganancia de peso de corderos

Para los datos de la Tabla 11.2.1, se obtiene $\bar{y} = 13$. Calculamos $SC(\text{total})$ como

$$\begin{aligned} SC(\text{total}) &= \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \\ &= [(8 - 13)^2 + (16 - 13)^2 + (9 - 13)^2] \\ &\quad + [(9 - 13)^2 + (16 - 13)^2 + (21 - 13)^2 + (11 - 13)^2 + (18 - 13)^2] \\ &\quad + [(15 - 13)^2 + (10 - 13)^2 + (17 - 13)^2 + (6 - 13)^2] \\ &= 246 \end{aligned}$$

Para estos datos, se obtiene $SC(\text{entre}) = 36$ y $SC(\text{intra}) = 210$. Podemos verificar que

$$246 = 36 + 210$$

Obtenemos también que $gl(\text{intra}) = 9$ y $gl(\text{entre}) = 2$. Podemos verificar que

$$gl(\text{total}) = 12 - 1 = 11 = 9 + 2$$

**La tabla del ANOVA**

Cuando se trabaja con los valores del ANOVA, es costumbre disponerlos en forma de tabla. El ejemplo siguiente muestra el formato típico de la tabla del ANOVA.

Ejemplo 11.2.6 Ganancia de peso de corderos

La Tabla 11.2.3 muestra el ANOVA para los datos de ganancia de peso de corderos. Nótese que la tabla del ANOVA muestra claramente la aditividad de la suma de cuadrados y de los grados de libertad.



Tabla 11.2.3 Tabla del ANOVA para las ganancias de peso de corderos			
Fuente	gl	SC	MC
Entre dietas	2	36	18,00
Intra dietas	9	210	23,33
Total	11	246	

Comentarios sobre terminología Aunque los términos «entre grupos» y «intra grupos» no son términos técnicos, son útiles para describir y comprender el modelo de ANOVA. El *software* de computador y otros textos se refieren comúnmente a esas fuentes de variabilidad como **tratamientos** (entre grupos) y **error** (dentro de los grupos).

Resumen de fórmulas

Para disponer de una referencia conveniente, presentamos en el recuadro las fórmulas de definición de los valores básicos del ANOVA.

Valores del ANOVA con fórmulas			
Fuente	gl	SC (suma de cuadrados)	CM (cuadrados medios)
Entre grupos	$I - 1$	$\sum_{i=1}^I n_i (\bar{y}_i - \bar{y})^2$	SC/gl
Dentro de los grupos	$n - I$	$\sum_{i=1}^I (n_i - 1) s_i^2$	SC/gl
Total	$n - 1$	$\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	

Ejercicios 11.2.1-11.2.7

11.2.1 La tabla que acompaña a este ejercicio muestra datos ficticios de tres muestras.

	Muestra		
	1	2	3
	48	40	39
	39	48	30
	42	44	32
	43		35
Media	43,00	44,00	34,00
DT	3,74	4,00	3,92

- Calcule SC(entre) y SC(intra).
- Calcule SC(total) y verifique la relación entre SC(entre), SC(intra) y SC(total).
- Calcule CM(entre), CM(intra) y s_{conjunta} .

11.2.2 Proceda como en el Ejercicio 11.2.1 con los siguientes datos:

	Muestra		
	1	2	3
	23	18	20
	29	12	16
	25	15	17
	23		23
			19
Media	25,00	15,00	19,00
DT	2,83	3,00	3,16

11.2.3 Para los siguientes datos, $SC(\text{intra}) = 116$ y $SC(\text{total}) = 338,769$.

Muestra		
1	2	3
31	30	39
34	26	45
39	35	39
32	29	37
	30	

- (a) Calcule $SC(\text{entre})$.
 (b) Calcule $CM(\text{entre})$, $CM(\text{intra})$, y s_{conjunta} .

11.2.4 La siguiente tabla de ANOVA está solo parcialmente completada.

Fuente	GL	SC	CM
Entre grupos	3		45
Dentro de los grupos	12	337	
Total		472	

- (a) Complete la tabla.
 (b) ¿Cuántos grupos había en el estudio?
 (c) ¿Cuántas observaciones había en total en el estudio?

11.2.5 La siguiente tabla de ANOVA está solo parcialmente completada.

Fuente	GL	SC	CM
Entre grupos	4		
Dentro de los grupos		964	
Total	53	1.123	

- (a) Complete la tabla.
 (b) ¿Cuántos grupos había en el estudio?
 (c) ¿Cuántas observaciones había en total en el estudio?

11.2.6 La siguiente tabla de ANOVA está solo parcialmente completada.

Fuente	GL	SC	CM
Entre grupos		258	
Dentro de los grupos	26		
Total	29	898	

- (a) Complete la tabla.
 (b) ¿Cuántos grupos había en el estudio?
 (c) ¿Cuántas observaciones había en total en el estudio?

11.2.7 Invente ejemplos de datos con

- (a) $SC(\text{entre}) = 0$ y $SC(\text{intra}) > 0$
 (b) $SC(\text{entre}) > 0$ y $SC(\text{intra}) = 0$
 (c) Para cada ejemplo, utilice tres muestras, cada una de ellas de tamaño 5.

11.3 Análisis del modelo ANOVA

En la Sección 11.2 presentamos la notación y_{ij} para la observación j en el grupo i . Suponemos que y_{ij} es una observación aleatoria del grupo i , siendo μ_i la media poblacional del grupo i . Utilizamos el análisis de la varianza para investigar la hipótesis nula de que $\mu_1 = \mu_2 = \dots = \mu_I$. Puede ser útil pensar en el ANOVA en términos del siguiente modelo:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

En este modelo, μ representa la media poblacional total, es decir, la media poblacional cuando todos los grupos se combinan. Si la hipótesis nula es cierta, entonces μ es la media poblacional común. Si la hipótesis nula es falsa, entonces al menos alguno de los μ_i se diferencian de la media poblacional total μ .

El término τ_i representa el efecto del grupo i , es decir, la diferencia entre la media poblacional del grupo i , μ_i , y la media poblacional total μ (τ es la letra griega «tau»). Por tanto,

$$\tau_i = \mu_i - \mu$$

La hipótesis nula

$$H_0: \mu_1 = \mu_2 = \dots = \mu_I$$

es equivalente a

$$H_0: \tau_1 = \tau_2 = \dots = \tau_I = 0$$

Si H_0 es falsa, entonces al menos uno de los grupos se diferencia de los otros. Si τ_i es positivo, entonces las observaciones del grupo i tienden a ser mayores que el promedio global. Si τ_i es negativo, los datos del grupo i tienden a ser menores que el promedio global.

El término ε_{ij} en el modelo representa a un error aleatorio asociado con la observación j del grupo i . Por tanto, el modelo

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

se puede expresar con palabras de la siguiente forma:

$$\text{observación} = \text{promedio global} + \text{efecto de los grupos} + \text{error aleatorio}$$

Para estimar el promedio global, μ , utilizamos la media global de los datos:

$$\hat{\mu} = \bar{y}$$

Asimismo, estimamos el promedio poblacional del grupo i mediante el promedio muestral del grupo i :

$$\hat{\mu}_i = \bar{y}_i$$

Como el efecto de los grupos es

$$\tau_i = \mu_i - \mu$$

estimamos τ_i como

$$\hat{\tau}_i = \bar{y}_i - \bar{y}$$

Finalmente, estimamos el error aleatorio, ε_{ij} , para la observación y_{ij} como

$$\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_i$$

Uniendo estos estimadores, tenemos

$$y_{ij} = \bar{y} + (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$$

o

$$y_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\varepsilon}_{ij}$$

Nota: Algunos autores utilizan la terminología SC(error) para lo que hemos denominado SC(intra). Esto es debido al hecho de que la componente dentro de los grupos $y_{ij} - \bar{y}_i$ estima el error aleatorio del modelo de ANOVA.

Ejemplo 11.3.1 Ganancia de peso de corderos

Para los datos del Ejemplo 11.2.1, la estimación de la media poblacional global es $\hat{\mu} = 13$. El efecto de los grupos estimado es

$$\hat{\tau}_1 = \bar{y}_1 - \bar{y} = 11 - 13 = -2$$

$$\hat{\tau}_2 = 15 - 13 = 2$$

y

$$\hat{\tau}_3 = 12 - 13 = -1$$

Por tanto, estimamos que la Dieta 2 aumenta la ganancia de peso en promedio en 2 lb (cuando se compara con el promedio de las tres dietas). La Dieta 1 disminuye la ganancia de peso en un promedio de 2 lb, y la Dieta 3 disminuye la ganancia de peso en 1 lb, en promedio. ■

Cuando se realiza un análisis de la varianza, estamos comparando los tamaños de los efectos muestrales en los grupos, los valores de τ_i , con los tamaños de los errores aleatorios en los datos, los valores de ε_{ij} . Podemos ver que

$$SC(\text{entre}) = \sum_{i=1}^I n_i \hat{\tau}_i^2$$

y

$$SC(\text{intra}) = \sum_{i=1}^I \sum_{j=1}^{n_i} \hat{\varepsilon}_{ij}^2$$

11.4 El test F global

La hipótesis nula global es

$$H_0: \mu_1 = \mu_2 = \dots = \mu_I$$

Consideremos el contraste de H_0 frente a la hipótesis alternativa no direccional (o omnidireccional)

$$H_A: \text{los } \mu_i \text{ no son todos iguales}$$

Nótese que H_0 es compuesta (a menos que $I = 2$) y, por tanto, el rechazo de H_0 no especifica cuáles de las μ_i son diferentes. Si rechazamos H_0 , podemos realizar después un análisis para efectuar comparaciones detalladas entre las μ_i . Contrastar la hipótesis nula global se puede comparar con mirar en un microscopio con una lente de pocos aumentos para ver si hay algo. Si encontramos algo, cambiamos a una lente con más aumentos para examinar la estructura con más detalle.

La distribución F

Las **distribuciones F** , denominadas así por el estadístico y genetista R. A. Fisher, son distribuciones de probabilidad que se usan en muchos tipos de análisis estadístico. La forma de una distribución F depende de dos parámetros: los **grados de libertad del numerador** y los **grados de libertad del denominador**. La Figura 11.4.1 muestra una distribución F con 4 grados de libertad en el numerador y 20 grados de libertad del denominador. Los valores críticos de la distribución F se dan en la Tabla 10 al final del libro. Nótese que la Tabla 10 ocupa 10 páginas, y cada página tiene un valor diferente de los grados de libertad del numerador. Como un ejemplo concreto, para 4 grados de libertad en el numerador y 20 grados de libertad en el denominador obtenemos en la Tabla 10 que $F(4,20)_{0,05} = 2,87$. Este valor se muestra en la Figura 11.4.1.

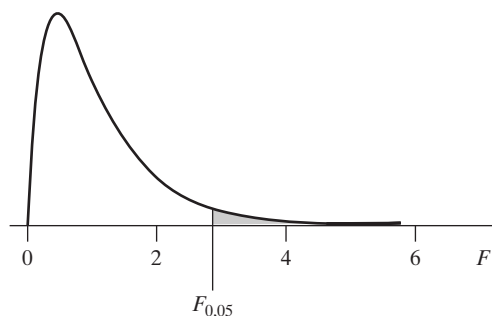


Figura 11.4.1 La distribución F con 4 grados de libertad en el numerador y 20 grados de libertad en el denominador

El test F

El **test F** es un test clásico de la hipótesis nula global. El estadístico de contraste, denominado **estadístico F** , se calcula como sigue:

$$F_s = \frac{\text{CM(entre)}}{\text{CM(intra)}}$$

A partir de las definiciones de los cuadrados medios (Sección 11.2) es claro que F_s será grande si las discrepancias entre las medias de los grupos ($\bar{Y}_i$) son grandes con respecto a la variabilidad dentro de los grupos. Por tanto, valores grandes de F_s tienden a proporcionar evidencia contra H_0 , es decir, evidencia a favor de una diferencia entre las medias de los grupos.

Para realizar el test F de la hipótesis nula global, los valores críticos se obtienen de una distribución F (Tabla 10) con

$$\text{gl del numerador} = \text{gl(entre)}$$

y

$$\text{gl del denominador} = \text{gl(intra)}$$

Se puede demostrar que (cuando se cumplen las condiciones adecuadas para la validez) la distribución nula de F_s es una distribución F con gl como los mencionados anteriormente.

El siguiente ejemplo ilustra el test F global.

Ejemplo 11.4.1 Ganancia de peso de corderos

Para el experimento sobre alimentación de corderos del Ejemplo 11.2.1, la hipótesis nula global y la hipótesis alternativa se pueden plantear verbalmente como

H_0 : la ganancia media de peso es la misma con las tres dietas.

H_A : La ganancia media de peso no es la misma con las tres dietas,

o simbólicamente como

$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_A : las μ_i no son todas iguales

Vimos en la Figura 11.2.2 que las tres medias muestrales no se diferencian mucho cuando se comparan con la variabilidad dentro de los grupos, lo que no es una evidencia muy fuerte contra H_0 . Confirmemos esta impresión visual realizando el test F con $\alpha = 0,05$. A partir de la tabla del ANOVA (Tabla 11.2.3) encontramos

$$F_s = \frac{18,00}{23,33} = 0,77$$

Los grados de libertad se pueden leer también de la tabla del ANOVA:

$$\text{gl del numerador} = 2$$

$$\text{gl del denominador} = 9$$

De la Tabla 10 tenemos que $F(2,9)_{0,20} = 1,93$, por lo que $P > 0,20$. Por tanto, hay una falta de evidencia significativa contra H_0 . Hay insuficiente evidencia para concluir que existe alguna diferencia entre las dietas con respecto a la ganancia media de peso poblacional. La diferencia observada en las ganancias medias de las muestras se puede atribuir fácilmente a la variación por azar. Debido a que este estudio es un experimento (y no un estudio observacional), podemos realizar incluso una afirmación algo más fuerte de los resultados: hay insuficiente evidencia para concluir que entre las tres dietas, alguna *afecta* a la ganancia de peso. ■

Relación entre el test F y el test t

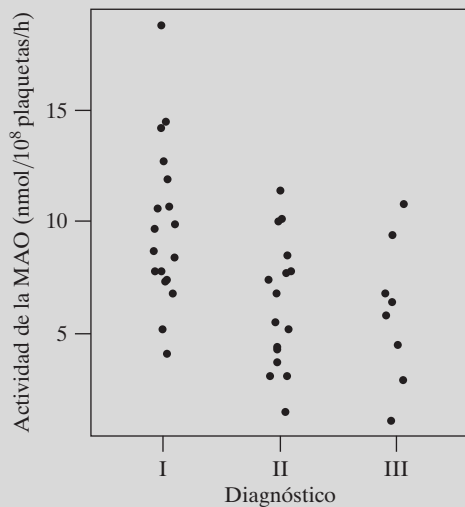
Supongamos que solo se van a comparar dos grupos ($I = 2$). Entonces se podría contrastar $H_0: \mu_1 = \mu_2$ contra $H_A: \mu_1 \neq \mu_2$ utilizando el test F o el test t . El test t del Capítulo 7 se puede modificar ligeramente sustituyendo la desviación típica de cada muestra por s_{conjunta} , definida en la Sección 11.2, antes de calcular el error típico de $(\bar{Y}_1 - \bar{Y}_2)$. Se puede demostrar que el test F y este test t «conjunto» son procedimientos realmente equivalentes. La relación entre los estadístico de contraste es $t_s^2 = F_s$. Es decir, el valor del estadístico F para cualquier conjunto de datos es necesariamente igual al cuadrado del valor del estadístico t (conjunto). La correspondiente relación entre los valores críticos es $t_{0,025}^2 = F_{0,05}$, $t_{0,005}^2 = F_{0,01}$, y así sucesivamente. Por ejemplo, supongamos $n_1 = 10$ y $n_2 = 7$. Entonces la distribución t apropiada tiene $\text{gl} = n_1 + n_2 - 2 = 15$, y $t_{15,0,025} = 2,131$, mientras que la distribución F tiene $\text{gl del numerador} = I - 1 = 1$ y $\text{gl del denominador} = n - I = 15$, por lo que $F(1,15)_{0,05} = 4,54$. Nótese que $(2,131)^2 = 4,54$. Debido a la equivalencia de los contrastes, la aplicación del test F para comparar las medias de dos muestras siempre dará exactamente el mismo P valor que el test t conjunto aplicado a los mismos datos.

Ejercicios 11.4.1-11.4.7

11.4.1 La monoamina oxidasa (MAO) es una enzima que se piensa juega un papel en la regulación del comportamiento. Para ver si diferentes categorías de pacientes con esquizofrenia tienen diferentes niveles de actividad de la MAO, unos investigadores extrajeron especímenes de sangre de 42 pacientes y midieron la actividad de la MAO en las plaquetas. Los resultados se resumen en la tabla que acompaña a este ejercicio. (Los valores expresan como nmol de producto de benzaldehído/ 10^8 plaquetas/hora)³. Los cálculos basados en los datos permiten obtener $SC(entre) = 136,12$ y $SC(intra) = 418,25$.

Diagnóstico	Actividad de la MAO		
	Media	DT	N.º de pacientes
Esquizofrenia crónica no diferenciada	9,81	3,62	18
No diferenciada con características paranoicas	6,28	2,88	16
Esquizofrenia paranoica	5,97	3,19	8

- (a) Se presentan a continuación diagramas de puntos de los datos. Basándose en esta gráfica, ¿parece que la hipótesis nula es cierta? ¿Por qué o por qué no?
- (b) Construya la tabla del ANOVA y contraste la hipótesis nula global con $\alpha = 0,05$.
- (c) Calcule la desviación típica conjunta, $s_{conjunta}$.



11.4.2 Se piensa que el estrés puede aumentar la susceptibilidad a las enfermedades debido a una supresión del sistema inmune. En un experimento para investigar esta teoría, se asignaron aleatoriamente 48 ratas a cuatro grupos de tratamiento: sin estrés, estrés débil, estrés moderado y estrés fuerte. Las condiciones de estrés se basaban en varios tipos de restricciones y choques eléctricos. Se midió en cada rata la concentración de linfocitos (células/ml $\times 10^{-6}$) en la sangre periférica con los resultados que se muestran en la tabla que acompaña a este ejercicio⁴. Los cálculos basados en los datos permiten obtener $SC(entre) = 89,036$ y $SC(intra) = 340,24$.

	Sin estrés	Estrés suave	Estrés moderado	Estrés alto
$\bar{y}$	6,64	4,84	3,98	2,92
s	2,77	2,42	3,91	1,45
n	12	12	12	12

- (a) Construya la tabla de ANOVA y contraste la hipótesis nula global con $\alpha = 0,05$.
- (b) Calcule la desviación típica conjunta, $s_{conjunta}$.

11.4.3 La beta endorfina humana (BEH) es una hormona segregada por la glándula pituitaria bajo condiciones de estrés. Un fisiólogo del ejercicio midió las concentraciones en sangre en reposo (sin estrés) de BEH en tres grupos de hombres: 15 que acababan de incorporarse a un programa de ejercicio físico, 11 que habían estado corriendo regularmente durante algún tiempo y 10 personas sedentarias. Los niveles de BEH (pg/ml) se muestran en la tabla siguiente⁵. Los cálculos basados en estos datos permiten obtener $SC(entre) = 240,69$ y $SC(intra) = 6.887,6$.

	Empezando programa de ejercicios	Corredores	Sedentarios
Media	38,7	35,7	42,5
DT	16,1	13,4	12,8
n	15	11	10

- (a) Plantee con palabras la hipótesis nula apropiada, en el contexto de este ejercicio.
- (b) Plantee la hipótesis nula con símbolos.
- (c) Construya la tabla del ANOVA y contraste la hipótesis nula. Utilice $\alpha = 0,05$.
- (d) Calcule la desviación típica conjunta, $s_{conjunta}$.

11.4.4 Se realizó un experimento en el que la medicación antiviral zanamivir se administra a pacientes que tenían la gripe. El tiempo hasta el alivio de los síntomas principales de la gripe se midió en tres grupos: 85 pacientes que habían

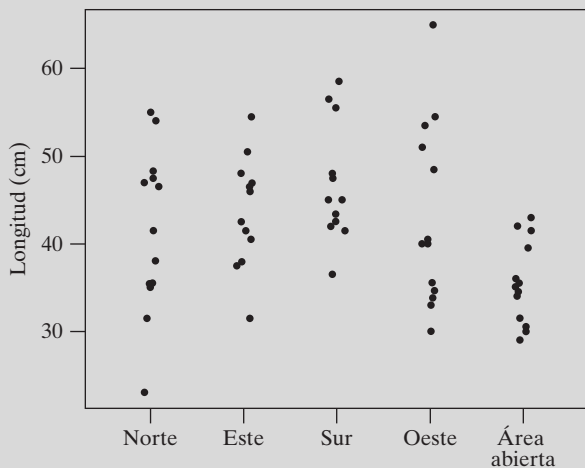
inhalado zanamivir, 88 pacientes que habían recibido zanamivir inhalado e intranasal, y 89 pacientes que recibieron un placebo. Los estadísticos resumen se muestran en la siguiente tabla⁶. Los valores del ANOVA son $SC(entre) = 53,67$ y $SC(intra) = 2.034,52$.

	Zanamivir inhalado	Zanamivir inhalado e intranasal	Placebo
Media	5,4	5,3	6,3
DT	2,7	2,8	2,9
n	85	88	89

- Plantee con palabras la hipótesis nula apropiada, en el contexto de este ejercicio.
- Plantee la hipótesis nula con símbolos.
- Construya la tabla del ANOVA y contraste la hipótesis nula. Utilice $\alpha = 0,05$.
- Calcule la desviación típica conjunta, $s_{conjunta}$.

11.4.5 Una investigadora recogió narcisos de los cuatro lados de un edificio y de un área abierta cercana. Deseaba saber si la longitud promedio del tallo de los narcisos dependía del lado del edificio en el que estaban creciendo. En la tabla siguiente se muestran los estadísticos resumen⁷. Los valores del ANOVA son $SC(entre) = 871,408$ y $SC(intra) = 3.588,54$.

	Norte	Este	Sur	Oeste	Abierta
Media	41,4	43,8	46,5	43,2	35,5
DT	9,3	6,1	6,6	10,4	4,7
n	13	13	13	13	13

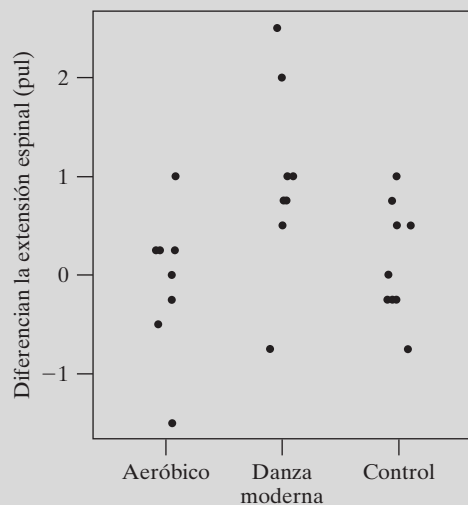


- Se presentan a continuación los diagramas de puntos de estos datos. Basándose en estos diagramas de puntos, ¿parece que la hipótesis nula es cierta? ¿Por qué o por qué no?
- Plantee la hipótesis nula en símbolos.
- Construya la tabla del ANOVA y contraste la hipótesis nula. Utilice $\alpha = 0,10$.

11.4.6 Una investigadora estudió la flexibilidad de 10 mujeres en una clase de ejercicio aeróbico, de 10 mujeres en una clase de danza moderna y de un grupo de control de 9 mujeres. Una medida que tomó de cada mujer fue la extensión espinal (es una medida de cuánto puede una mujer curvar su espalda). Se realizaron medidas antes y después de un periodo de entrenamiento de 16 semanas. Se midió en cada mujer el cambio en la extensión espinal. Los estadísticos resumen se muestran en la tabla siguiente⁸. Los valores del ANOVA son $SC(entre) = 7,04$ y $SC(intra) = 15,08$.

	Aeróbico	Danza moderna	Control
Media	-0,18	0,98	0,13
DT	0,80	0,86	0,57
n	10	10	9

- En la figura de abajo se muestran diagramas de puntos de estos datos. Basándose en los diagramas de puntos, ¿parece que la hipótesis nula es cierta? ¿Por qué o por qué no?
- Plantee la hipótesis nula mediante símbolos.
- Construya en la tabla del ANOVA y contraste la hipótesis nula. Utilice $\alpha = 0,01$.



11.4.7 La siguiente salida de computador es de un análisis de varianza donde se compararon cosechas (por acre) de diferentes variedades de avena⁹.

Fuente	gl	Suma de cuadrados	Cuadrados medios	Razón F	Prob
Grupo	2	76,8950	38,4475	0,40245	0,6801
Error	9	859,808	95,5342		
Total	11	936,703			

- (a) ¿Cuántas variedades (grupos) hacían el experimento?
 (b) Plantee la conclusión del ANOVA.
 (c) ¿Cuál es la desviación típica conjunta, s_{conjunta} ?

11.5 Aplicabilidad de los métodos

Como todos los otros métodos de inferencia estadística, los cálculos e interpretaciones del ANOVA se basan en ciertas condiciones.

Condiciones estándar

Las técnicas de ANOVA descritas en este capítulo, incluyendo el test F global, son válidas si se cumplen las siguientes condiciones.

1. *Condiciones de diseño*
 - (a) Debe ser razonable considerar los grupos de observaciones como muestras aleatorias de sus respectivas poblaciones.
 - (b) Las I muestras deben ser independientes entre sí.
2. *Condiciones sobre la población.* Las I distribuciones poblacionales deben ser (aproximadamente) normales con las mismas desviaciones típicas:

$$\sigma_1 = \sigma_2 = \dots = \sigma_I$$

Estas condiciones son extensiones de la condición dada en el Capítulo 7 para el test t de muestras independientes, con la condición añadida de que las desviaciones típicas deben ser iguales. La condición de poblaciones normales con la misma desviación típica es menos crucial si los tamaños muestrales (n_i) son grandes y aproximadamente iguales.

Verificación de las condiciones

Las condiciones de diseño se pueden verificar como en el test t de muestras independientes. Para comprobar la condición 1(a), se buscan sesgos o estructuras jerárquicas en el conjunto de los datos. Un diseño completamente aleatorizado asegura la independencia de las muestras [condición 1(b)]. Si las unidades se asignan a grupos de tratamiento de una forma no aleatoria (por ejemplo, mediante un diseño por bloques aleatorizados que se presentará en la Sección 11.6), o si las observaciones de la misma unidad experimental aparecen en diferentes muestras (por ejemplo, para $I = 2$, datos pareados, como se vio en el Capítulo 9), entonces las muestras no son independientes.

Como en el test t de muestras independientes, las condiciones de la población se pueden comprobar aproximadamente a partir de los datos. Para comprobar la normalidad, se pueden hacer histogramas o gráficas de probabilidad normal para cada muestra. Otra opción es hacer un único histograma o una gráfica de probabilidad normal de las desviaciones ($y_{ij} - \bar{y}_i$) de todas las muestras combinadas. En el contexto del análisis de la varianza, estas desviaciones de las medias de los grupos se denominan **residuos**. Por tanto, un residuo mide la distancia de un valor de los datos a su respectiva media de grupo.

La igualdad de las DT poblacionales se comprueba comparando las DT muestrales. Un truco fácil es dibujar las DT frente a las medias ($\bar{y}_i$) para buscar una tendencia. Otro planteamiento es hacer una gráfica de los residuos ($y_{ij} - \bar{y}_i$) frente a las medias ($\bar{y}_i$). Como regla práctica, deseáramos que las DT muestral máxima dividida por la DT muestral mínima sea aproximadamente menor que 2. Si esta razón es mucho mayor que 2, entonces no podemos confiar en el P valor del ANOVA, particularmente si los tamaños muestrales son pequeños y distintos. En particular, si los tamaños muestrales no son iguales y la DT muestral de una muestra es bastante mayor que las otras DT, entonces el P valor puede ser muy inexacto.

Ejemplo 11.5.1 Ganancia de peso de corderos

Considere el experimento sobre alimentación de corderos del Ejemplo 11.2.2. La Figura 11.2.1 (en la Sección 11.2) muestra que la variabilidad dentro de los grupos es casi igual en las tres dietas: las tres DT muestrales son 4,36, 4,95 y 4,97. La Figura 11.5.1 es una gráfica de probabilidad normal de los 12 residuos ($y_{ij} - \bar{y}_i$) (1 de la Dieta 1, 5 de la Dieta 2 y 4 de la Dieta 3). Esta gráfica es casi lineal, por lo que no hay evidencia que plantee dudas sobre la condición de normalidad. ■

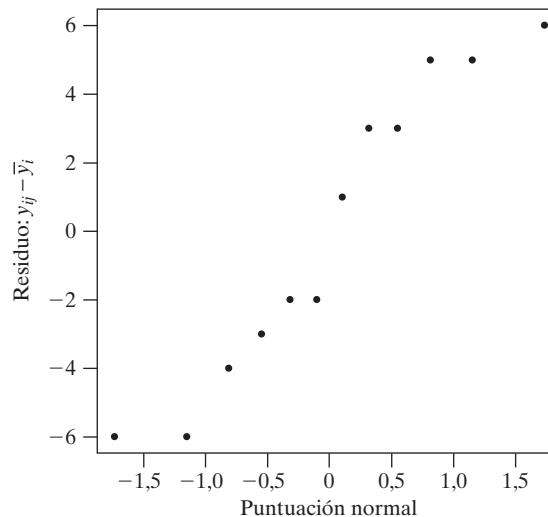


Figura 11.5.1 Gráficas de probabilidad normal de los residuos ($y_{ij} - \bar{y}_i$) de los datos de ganancia de peso

Ejemplo 11.5.2 Maíz dulce

Consideremos los datos de maíz dulce del Ejemplo 11.1.1. La Figura 11.5.2(a) muestra los datos, cada grupo con su propio símbolo. La Figura 11.5.2(b) muestra los residuos ($y_{ij} - \bar{y}_i$) frente a las medias de los grupos ($\bar{y}_i$) (que se conoce también como **valor ajustado** en el contexto de análisis de la varianza). Esta segunda gráfica muestra que la variabilidad (medida visualmente como extensión vertical) no cambia apreciablemente con los cambios en la media (eso es bueno, ya que si la variabilidad creciera al aumentar la media, entonces se violaría la condición 2). ■

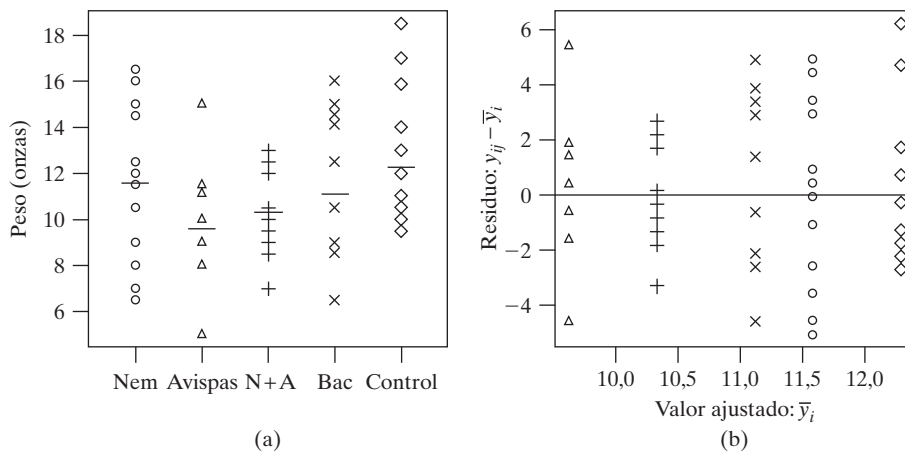


Figura 11.5.2 Gráfica de los residuos frente a la media muestral para los datos del maíz dulce

Aunque se podría observar una gráfica básica de los datos, como la de la Figura 11.5.2(a), para inspeccionar visualmente que las DT son similares en todos los grupos, al dibujar los datos como en la Figura 11.5.2(b) se obtienen más ventajas visuales. Primero, examinando los residuos (Figura 11.5.2(b)) y no los datos iniciales (Figura 11.5.2(a)), la gráfica se puede observar de izquierda a derecha, permitiendo que los ojos comparen visualmente de manera más clara la variabilidad entre los grupos sin ser distraídos por el cambio de las medias. Además, una violación común del requerimiento de DT iguales es que las DT de los grupos crezcan con las medias. Para ilustrar esta violación, considere los datos ficticios que se muestran en la Figura 11.5.3(a), formados por cinco grupos de tratamiento y siete observaciones por grupo. Claramente la variabilidad no es la misma en los cinco grupos. La gráfica de los residuos frente a las medias de la Figura 11.5.3(b) expone este problema más claramente y muestra que las DT (representadas por la dispersión vertical) aumentan con la media. Este efecto se describe comúnmente como residuos con forma de *embudo* o *cuerno*.

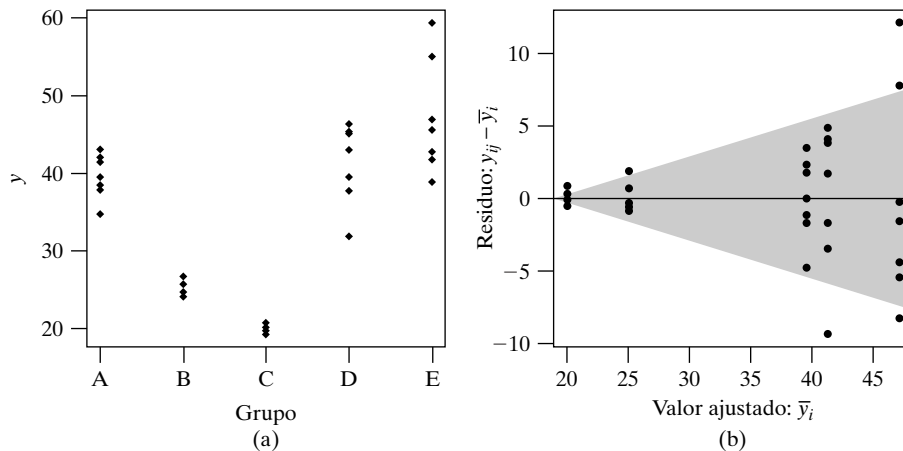


Figura 11.5.3 Gráfica de los residuos frente a las medias muestrales para un conjunto de datos ficticio en el que la desviación típica crece con la media

Ejemplo 11.5.3 Maíz dulce

Considerando de nuevo los datos de maíz dulce del Ejemplo 11.2.1, examinemos la normalidad de los grupos mediante el examen de los residuos. La Figura 11.5.4 contiene un histograma y una gráfica de probabilidad normal de los 60 residuos ($y_{ij} - \bar{y}_i$). La forma acampanada de la gráfica (a) y la linealidad de la gráfica (b) dejan pocas dudas sobre la condición de normalidad. ■

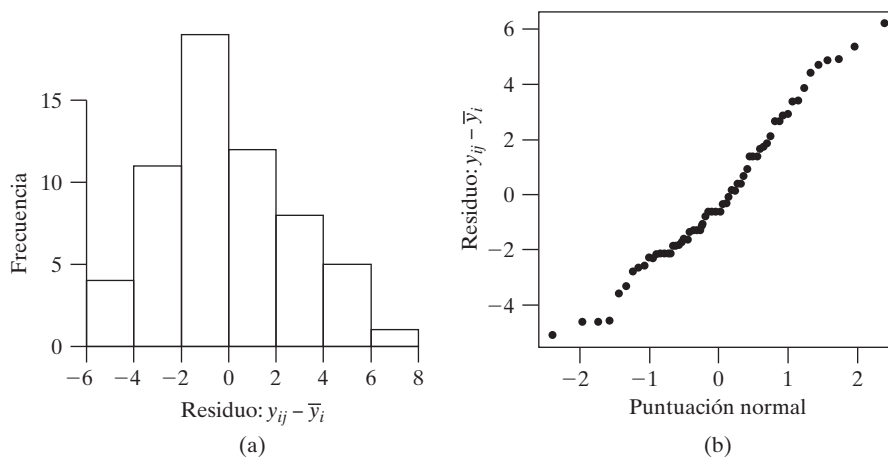


Figura 11.5.4 Histograma y gráfica de probabilidad normal de las desviaciones ($y_{ij} - \bar{y}_i$) en los datos de maíz dulce

Análisis posterior

Además de su relevancia para el test F , las condiciones estándar son necesarias en muchos métodos clásicos para un análisis posterior de los datos.

Si las I poblaciones tienen la misma DT, entonces una estimación conjunta de esa DT a partir de los datos es

$$s_{\text{conjunta}} = \sqrt{\text{CM}(\text{intra})}$$

del ANOVA. Esta desviación típica conjunta s_{conjunta} es un estimador mejor que cualquier DT muestral individual, ya que s_{conjunta} está basada en más observaciones.

Una forma simple de ver la ventaja de s_{conjunta} es considerar el error típico de una media muestral individual (Y_i), que se puede calcular como

$$\text{ET}_{\bar{Y}_i} = \frac{s_{\text{conjunta}}}{\sqrt{n_i}}$$

siendo n_i el tamaño de la muestra individual. Los gl asociados con este error típico son $\text{gl}(\text{intra})$, que es la suma de los grados de libertad de todas las muestras. Por el contrario, si se utilizara la DT individual al calcular $\text{ET}_{\bar{Y}_i}$, solo habría $(n_i - 1)$ gl. Cuando se utiliza el ET para inferencia, un valor mayor de gl produce valores críticos menores (véase la Tabla 4), que a su vez conducen a un aumento de la potencia y a intervalos de confianza más estrechos.

En las Secciones opcionales 11.7 y 11.8 consideraremos métodos para el análisis detallado de las medias de los grupos $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_I$. Como el test F , estos métodos se han diseñado para muestras independientes de poblaciones normales con desviaciones típicas iguales. Los métodos utilizan errores típicos basados en la estimación de la desviación típica conjunta s_{conjunta} .

Ejercicios 11.5.1-11.5.2

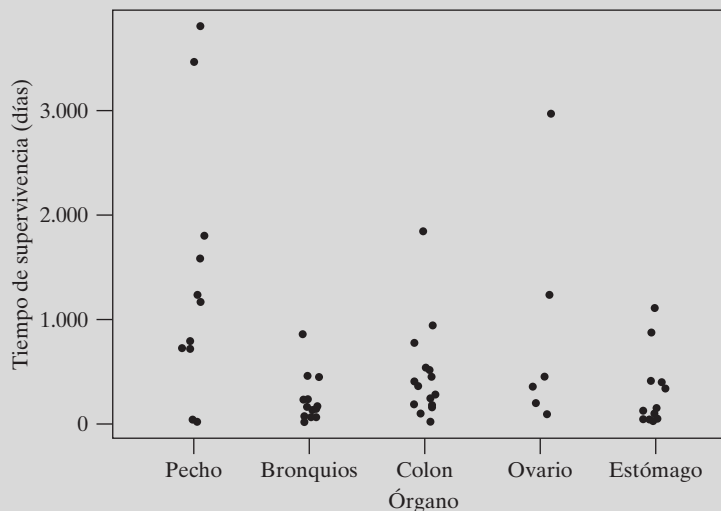
11.5.1 En referencia a los datos de linfocitos del Ejercicio 11.4.2, el test F global se basa en ciertas condiciones relativas a las distribuciones poblacionales.

- Plantee las condiciones.
- ¿Qué características de los datos sugieren que las condiciones podrían ser dudosas en este caso?

11.5.2 Una serie de pacientes con cánceres avanzados de estómago, bronquios, colon, ovario o pecho se trataron con

ascorbato. El propósito del estudio era determinar si los tiempos de supervivencia diferían con respecto al órgano afectado por el cáncer. La variable de interés es el tiempo de supervivencia (en días). Se presentan a continuación diagramas de puntos en paralelo de los datos iniciales¹⁰.

Se realizó un ANOVA después de modificar los datos iniciales mediante una transformación de raíz cuadrada. Había dos razones (relacionadas) para transformar los datos. ¿Cuáles eran esas dos razones?



11.6 Diseño de bloques aleatorizados con un factor

El diseño completamente aleatorizado no hace distinciones entre las unidades experimentales. Muchas veces se puede mejorar un experimento con un planteamiento más refinado, uno que aproveche los patrones conocidos de variabilidad en las unidades experimentales.

En un **diseño de bloques aleatorizados** se agrupan en primer lugar las unidades experimentales en conjuntos, o **bloques**, de unidades relativamente similares y después se asignan aleatoriamente los tratamientos a cada bloque. He aquí un ejemplo.

Ejemplo 11.6.1 Alfalfa y lluvia ácida

Unos investigadores estaban interesados en el efecto que tenía el ácido en la velocidad de crecimiento de plantas de alfalfa. En un experimento, crearon tres grupos de tratamiento: ácido bajo, ácido alto y control. La variable de respuesta del experimento era la altura de las plantas de alfalfa en un recipiente de poliestireno tras cinco días de crecimiento*. Utilizaron cinco recipientes para cada uno de los tres tratamientos, con un total de 15 observaciones. Sin embargo, los recipientes se dispusieron cerca de una ventana ya que deseaban investigar el efecto de las diferentes cantidades de luz solar. Por tanto, crearon 5 bloques (cada bloque a una distancia fija de la ventana: el bloque 1 era el más cercano y el bloque 5 el más lejano). Dentro de cada bloque en los tres tratamientos se asignaron aleatoriamente, como muestra la Figura 11.6.1[†].

	Bloque 1	Bloque 2	Bloque 3	Bloque 4	Bloque 5
Ventana	alto	control	control	control	alto
	control	bajo	alto	bajo	bajo
	bajo	alto	bajo	alto	control

Organización de los bloques en el experimento de la alfalfa

Figura 11.6.1 Diseño del experimento de la alfalfa

El Ejemplo 11.6.1 es una ilustración de un diseño en bloques aleatorizados. Para realizar este tipo de diseño, el investigador crea o identifica bloques adecuados de unidades experimentales y seguidamente asigna aleatoriamente los tratamientos dentro de cada bloque, de forma que cada tratamiento aparezca en cada bloque[†]. En el Ejemplo 11.6.1, las filas de recipientes en cada una de las cinco distancias a la ventana sirven como bloques. En general, crearemos bloques para reducir o eliminar la variabilidad causada por variables externas, de forma que la precisión del experimento aumente. Es conveniente que las unidades experimentales dentro de un bloque sean homogéneas; deseamos que las variables extrañas ocurran *entre* los bloques. Presentamos a continuación más ejemplos de diseños en bloques aleatorizados en experimentos biológicos.

Ejemplo 11.6.2 Bloques por camadas

¿Cómo afecta la experiencia a la anatomía del cerebro? En un experimento típico para estudiar esta cuestión, se situaron ratas jóvenes en tres entornos durante 80 días:

T_1 : *Entorno estándar*. La rata se pone con una única compañera en una jaula de laboratorio estándar.

T_2 : *Entorno enriquecido*. La rata se pone con varias compañeras en una jaula grande, donde se ponen también varios objetos de juego.

T_3 : *Entorno empobrecido*. La rata vive sola en una jaula estándar.

Al final de los 80 días del experimento se realizaron varias medidas anatómicas en los cerebros de las ratas.

* De forma más precisa, la variable de respuesta era la altura media de las plantas de un recipiente, por lo que la unidad observacional era el recipiente, y no las plantas individuales.

[†] Estrictamente hablando, el diseño que presentamos se denomina *diseño de bloques completos aleatorizados* porque cada tratamiento aparece en cada bloque. En un *diseño de bloques incompletos*, cada bloque contiene algunos de los tratamientos, pero no necesariamente todos.

Supongamos que un investigador planea realizar el experimento anterior utilizando 30 ratas. Para minimizar la variación en la respuesta, los 30 animales eran machos, de la misma edad y raza. Para reducir aún más la variación, el investigador puede aprovechar la similitud de los animales de la misma camada. En este planteamiento, el investigador tendría tres ratas macho de cada una de 10 camadas. Los tres machos de cada camada se asignarían aleatoriamente: uno a T_1 , uno a T_2 y el otro a T_3 ¹². ■

Otra forma de visualizar el diseño experimental es mediante una tabla, como se muestra en la Tabla 11.6.1. Cada «Y» de la tabla representa una observación en una rata. Utilizando la disposición de la Tabla 11.6.1, el experimentador puede comparar las respuestas de las ratas que recibieron tratamientos *diferentes* pero que son de la *misma* camada. Esas comparaciones no están afectadas por cualquier diferencia (genética ni de otro tipo) que pueda existir entre una camada y otra.

Tabla 11.6.1 Formato de los datos de cerebros de ratas			
	Tratamiento		
	T_1	T_2	T_3
Camada 1	Y	Y	Y
Camada 2	Y	Y	Y
Camada 3	Y	Y	Y
⋮	⋮	⋮	⋮
Camada 10	Y	Y	Y

Ejemplo 11.6.3 Bloques dentro de sujetos (emparejamiento)

Un dermatólogo está planificando un estudio para comparar dos lociones médicas en cuanto a su efectividad para tratar el acné. En el estudio van a participar 20 pacientes. Cada paciente utilizará la loción A en un lado de su cara y la loción B en el otro. El dermatólogo observará la mejora en cada lado durante un periodo de tres meses. En cada paciente, el lado de la cara que va a recibir la loción A se selecciona aleatoriamente y el otro lado recibe la loción B. Las botellas del medicamento tienen etiquetas codificadas de forma que ni los pacientes ni los médicos conocen qué botella contiene A y qué botella contiene en B (es decir, además de utilizar bloques el experimento también es ciego)¹³. Este ejemplo, con bloques de tamaño 2, es un ejemplo de emparejamiento. El lado izquierdo de la cara está emparejado con el lado derecho. En el Capítulo 8 hemos considerado el análisis de datos pareados. ■

Ejemplo 11.6.4 Bloques en un estudio agrícola de campo

Cuando se comparan varias variedades de grano, los agrónomos generalmente plantan muchas parcelas de campo para cada variedad y miden la cosecha de cada parcela. Las diferencias en las cosechas pueden reflejar no solamente diferencias genuinas entre las variedades, sino también diferencias entre las parcelas en cuanto a fertilidad del suelo, pH, capacidad de almacenamiento de agua... En consecuencia, la disposición espacial de las parcelas en el campo es importante. Una forma eficiente de utilizar el área de campo disponible es dividir dicha área en regiones grandes (los bloques) y subdividir cada bloque en varias parcelas. Dentro de cada bloque se asignan aleatoriamente las diversas variedades de grano a las parcelas, de forma que se realiza una aleatorización separada en cada bloque. Por ejemplo, supongamos que deseamos probar cuatro variedades de cebada. Entonces, cada bloque podría contener cuatro parcelas. La asignación aleatorizada resultante podría ser como muestra la Figura 11.6.2, que es un mapa esquemático del campo. Los «tratamientos» T_1 , T_2 , T_3 y T_4 son las cuatro variedades de cebada. ■

Creación de los bloques

Como muestran los ejemplos anteriores, los bloques son una forma de *organizar* la variación inherente que existe entre las unidades experimentales. Idealmente, los bloques se deben disponer de forma que incrementen la información disponible en el experimento. Para conseguir este objetivo, *el investigador debe intentar crear bloques que sean tan*

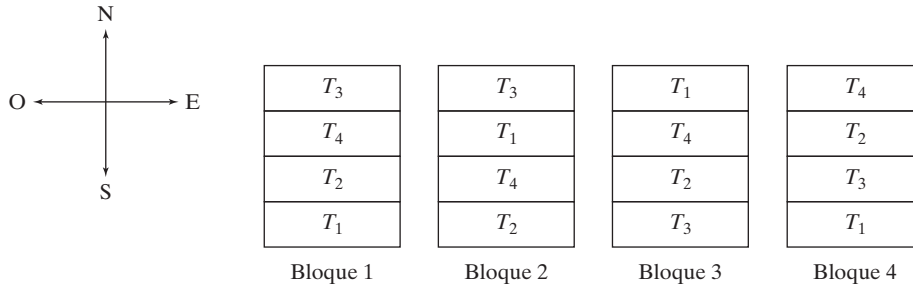


Figura 11.6.2 Disposición de un diseño agrícola en bloques aleatorizados

homogéneos dentro de sí mismos como sea posible, de forma que la variación inherente entre unidades experimentales se convierta, tanto como sea posible, en variación entre bloques en vez de variación dentro de los bloques. Este principio se ilustró en los ejemplos anteriores (por ejemplo, en el Ejemplo 11.6.2, donde los bloques por camada explotaron el hecho de que los machos de la misma camada son más similares entre sí que los de camadas distintas). A continuación sigue otra ilustración.

Ejemplo 11.6.5 Estudio de campo agrícola

En el experimento de la cebada del Ejemplo 11.6.4, ¿cómo determinarían los agrónomos la mejor disposición o diseño de los bloques en el campo? Diseñarían los bloques para aprovechar cualquier conocimiento previo que pudieran tener de los modelos de fertilidad del campo. Por ejemplo, si supieran que existe un gradiente de fertilidad este-oeste en el campo (quizá porque el campo está en pendiente de este a oeste, con el resultado de que el extremo oeste tiene una capa más gruesa de suelo bueno o recibe una mejor irrigación), entonces podrían elegir bloques como los de la Figura 11.6.2. El diseño maximiza las diferencias de suelo entre los bloques y minimiza las diferencias entre las parcelas dentro de cada bloque. (Pero incluso si un campo parece ser uniforme, los bloques se utilizan habitualmente en experimentos agrícolas, porque las parcelas que están juntas en el campo son generalmente más similares que las que están alejadas).

Para aumentar la solidez de este ejemplo, tomemos un conjunto de datos del experimento de bloques aleatorizados de la cebada. Cada valor de la Tabla 11.6.2 muestra la cosecha (celemines de cebada por acre) de una parcela de 3,5 pies de ancho por 80 pies de largo¹⁴.

	Bloque 1	Bloque 2	Bloque 3	Bloque 4	Media de la variedad
Variedad 1	93,5	66,6	50,5	42,4	63,3
Variedad 2	102,9	53,2	47,4	43,8	61,8
Variedad 3	67,0	54,7	50,0	40,1	53,0
Variedad 4	86,3	61,3	50,7	46,4	61,2
Media de bloque	87,4	59,0	49,7	43,2	

Al observar la Tabla 11.6.2, parece que el potencial de cosecha de los bloques varía grandemente. Los datos indican un gradiente definido de fertilidad del bloque 1 al bloque 4. Debido al diseño de bloques, la comparación de las variedades queda relativamente afectada por el gradiente de fertilidad. Por supuesto, también parece haber variación sustancial dentro de los bloques. [Podría ser un ejercicio interesante examinar detenidamente los datos y preguntarnos si las diferen-

cias observadas entre variedades son lo suficientemente grandes como para concluir que, por ejemplo, la variedad 1 es superior (en cosecha media) a la variedad 3, utilizando nuestra intuición en vez de un análisis estadístico formal. La verdad se revela en la Nota 14].

El procedimiento de aleatorización

Una vez se han creado los bloques, la asignación de unidades experimentales a los bloques es directa: es como si se realizara un miniexperimento dentro de cada bloque. La aleatorización se realiza separadamente para cada bloque, como se ilustra en el siguiente ejemplo.

Ejemplo 11.6.6 Estudio de campo agrícola

Consideremos el experimento agrícola de campo del Ejemplo 11.6.4. En el bloque 1, etiquetaremos las parcelas como 1, 2, 3, 4, de norte a sur (véase la Figura 11.6.2). Asignaremos una variedad a cada parcela. La asignación procede según el diseño completamente aleatorizado, escogiendo parcelas aleatoriamente entre las cuatro, y asignando la primera parcela elegida a T_1 , la segunda a T_2 , y así sucesivamente. Por ejemplo, utilizando un computador para permutar aleatoriamente los números 1 a 4 (o incluso cartas previamente barajadas numeradas de 1 a 4) podríamos obtener la secuencia 4, 3, 1, 2 que nos llevaría a la siguiente asignación de tratamientos.

Bloque 1

- T_1 : Parcela 4
- T_2 : Parcela 3
- T_3 : Parcela 1
- T_4 : Parcela 2

Esta es de hecho la asignación que se muestran la Figura 11.6.2 para el bloque 1. Podemos repetir así este procedimiento para los bloques 2, 3, y así sucesivamente. ■

Análisis de datos de un experimento de bloques aleatorizados

De la misma forma que no podemos utilizar un test t de dos muestras cuando los datos están emparejados, cuando en un experimento se han hecho bloques, no podemos utilizar los métodos de ANOVA de la Sección 11.4. En su lugar, utilizaremos un modelo de **ANOVA de bloques aleatorizados**. Ilustraremos los conceptos reconsiderando el experimento de la alfalfa y la lluvia ácida del Ejemplo 11.6.1 en el que los investigadores diseñaron los bloques del experimento basándose en filas de recipientes situadas paralelas a una ventana, de forma que cada bloque tuviera aproximadamente la misma exposición a la luz. Los datos se muestran en la Tabla 11.6.3 y se representan gráficamente en la Figura 11.6.3.

Tabla 11.6.3 Altura de la planta de alfalfa después de cinco días (cm)				
	Ácido alto	Ácido bajo	Control	Media de bloque
Bloque 1	1,30	1,78	2,67	1,917
Bloque 2	1,15	1,25	2,25	1,550
Bloque 3	0,50	1,27	1,46	1,077
Bloque 4	0,30	0,55	1,66	0,837
Bloque 5	1,30	0,80	0,80	0,967
Media de tratamiento = $\bar{y}_i$	0,910	1,130	1,768	
n	5	5	5	

Nuestra hipótesis nula habitual del ANOVA para comparar I poblaciones o tratamientos es

$$H_0: \mu_1 = \mu_2 = \dots = \mu_I$$

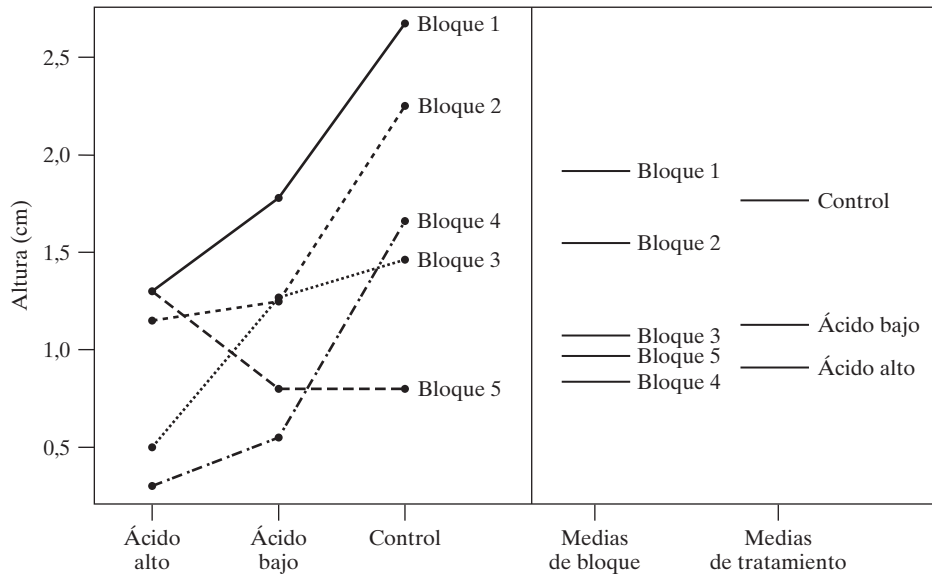


Figura 11.6.3 Diagramas de puntos de los datos de crecimiento de alfalfa con un resumen de las medias de bloque y de tratamiento

Ejemplo 11.6.7 Alfalfa y lluvia ácida

La hipótesis nula en el experimento de crecimiento de alfalfa es que el ácido no tiene efecto en el crecimiento durante cinco días. (Podemos hacer una afirmación causal fuerte como esta porque se trata de un experimento). De forma más directa, la hipótesis nula es que el crecimiento medio en cinco días es el mismo para los tres tratamientos (ácido alto, ácido bajo y control).

$$H_0: \mu_1 = \mu_2 = \mu_3$$

Esta hipótesis se puede contrastar con un test F de análisis de varianza, pero primero hay que eliminar la variabilidad de los datos debida a las diferencias entre bloques. Para hacer esto, ampliamos el modelo de ANOVA presentado en la Sección 11.3 planteando el modelo siguiente:

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk}$$

En este modelo, y_{ijk} es la k -ésima observación cuando se aplica el tratamiento i en el bloque j . (En el Ejemplo 11.6.1 solo había una observación de cada tratamiento en cada bloque, pero en general puede haber más de una). Aquí, como antes, μ representa la media poblacional global y el término τ_i representa el efecto del grupo i (es decir, el tratamiento i). El nuevo término del modelo es β_j , que representa el efecto del bloque j -ésimo.

Visualización de los efectos de bloque

Para visualizar cómo afectan los bloques a nuestro ANOVA, podemos ver nuestro modelo de una forma ligeramente diferente:

$$(y_{ijk} - \tau_i) = \mu + \beta_j + \varepsilon_{ijk}$$

El miembro izquierdo de la ecuación describe los datos después de que se han eliminado los efectos de los tratamientos. Utilizando nuestros datos, estimamos este miembro izquierdo como

$$y_{ijk} - \hat{\tau}_i = y_{ijk} - \bar{y}_{i.}$$

Es decir, dentro de cada grupo de tratamiento, se resta la media del tratamiento a cada valor de los datos*. Ya hemos visto esto antes: en el contexto del ANOVA de un factor (Sección 11.2) denominábamos a esas diferencias desviaciones o residuos. La Figura 11.6.4 es una gráfica de las desviaciones de las medias de los tratamientos para los datos de la alfalfa, divididas por bloques. Podemos ver que todavía hay mucha estructura en los datos: las desviaciones medias de los bloques 1 y 2 son mayores que cero, mientras que los bloques 3, 4 y 5 están por debajo de cero (lo que corresponde a un crecimiento medio alto cerca de la ventana y un crecimiento medio bajo lejos de la ventana). El hecho de que estas desviaciones medias no sean todas cero es una consecuencia de la variabilidad debida a los bloques. En breve describiremos cómo medir la variabilidad de estas desviaciones medias de los bloques a través de los **cuadrados medios por bloques**, o **CM(bloques)**.

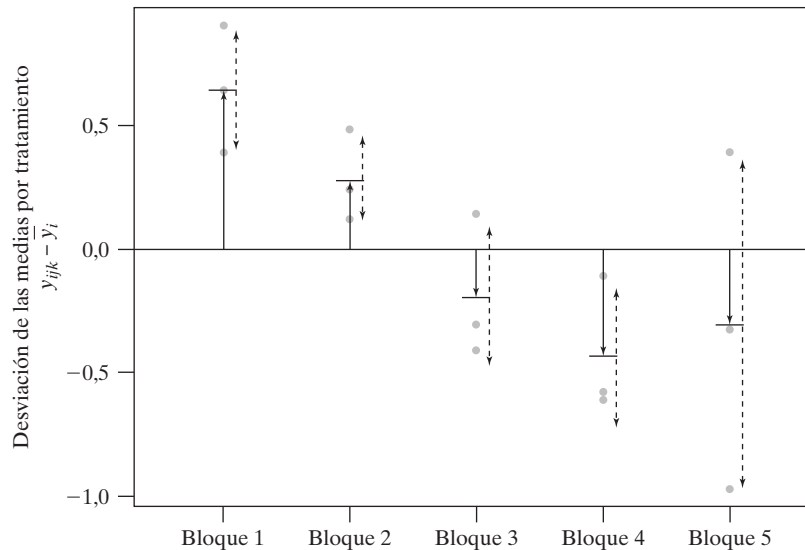


Figura 11.6.4 Desviaciones de las medias de los tratamientos para los datos de crecimiento de alfalfa, por bloques

Para visualizar cómo el tener en cuenta esta variación de bloque a bloque mejora nuestra capacidad para detectar los efectos de los tratamientos, consideremos los datos sobre la alfalfa y la lluvia ácida que se muestran en la Figura 11.6.5. La Figura 11.6.5(a) muestra los datos de crecimiento para cada grupo de tratamiento y, simplemente, ignora completamente los bloques mientras que la Figura 11.6.5(b) muestra los datos de crecimiento después de ajustar los efectos estimados de bloque†. Aunque la variabilidad entre las medias por tratamiento no cambia entre las parcelas, observamos que la variabilidad dentro de los grupos de tratamiento es mucho más pequeña después de tener en cuenta los bloques y, por tanto, las diferencias entre los tratamientos son más pronunciadas.

El test F para diseño de bloques completamente aleatorizados de un factor

Recuérdese que el test F del ANOVA es una razón que compara a la variabilidad entre las medias de los tratamientos con la variabilidad dentro de los grupos. Como se vio en la Figura 11.6.5, al tener en cuenta los bloques se reduce la variabilidad dentro de los grupos y, por tanto, se aumenta el valor del estadístico F . Presentaremos ahora brevemente los cálculos requeridos para realizar la tabla del ANOVA para el test F de diseño de bloques completamente aleatorizados.

* En este caso escribimos $\bar{y}_i$, en vez de $\bar{y}_i$ para distinguir las medias de los tratamientos de las medias de los bloques $\bar{y}_j$.

† Para tener en cuenta los bloques, los datos de crecimiento ajustados en el eje y para cada grupo de tratamiento se calculan como $y_{ijk} - \bar{y}_j$.

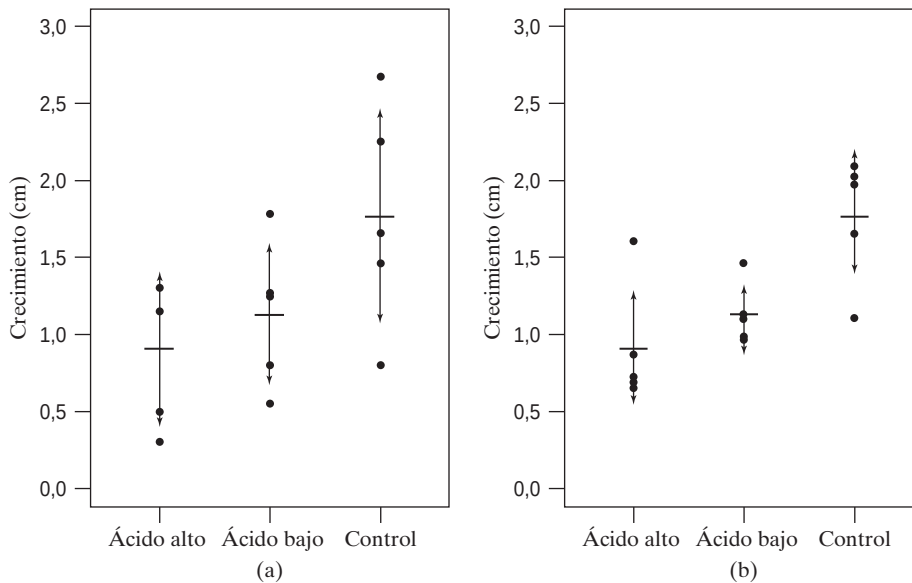


Figura 11.6.5 Visualización del efecto de los bloques cuando se compara el crecimiento medio bajo los tres tratamientos de ácido en el experimento de la alfalfa. La gráfica (a) muestra los datos iniciales de crecimiento mientras que (b) muestra los datos de crecimiento después de ajustar los efectos estimados de los bloques. Las medias por tratamiento se indican mediante líneas horizontales y las desviaciones típicas dentro de los grupos mediante flechas

En la Sección 11.2 sobre el ANOVA de un factor, presentamos cómo la suma de cuadrados total, $SC(\text{total})$ se divide en $SC(\text{entre})$, que mide la variabilidad atribuible a las diferencias entre las medias de los tratamientos, y $SC(\text{intra})$, que mide la variación aleatoria no explicada en los datos. En un experimento de bloques aleatorizados utilizamos $SC(\text{tratamientos})$ en vez de $SC(\text{entre})$ para describir la variabilidad entre las medias de los tratamientos, aclarando así que estamos viendo variabilidad entre tratamientos y no entre bloques. En un experimento de bloques aleatorizados dividiremos también el valor $SC(\text{intra})$ del ANOVA de un factor en dos partes: $SC(\text{bloques})$, que mide la variabilidad entre las medias de los bloques, y $SC(\text{intra})$, que mide la variación restante no explicada en los datos. Por tanto, tenemos

$$\begin{array}{ll} \text{ANOVA de un factor:} & SC(\text{total}) = SC(\text{intra}) + SC(\text{entre}) \\ & \swarrow \quad \searrow \\ \text{ANOVA de un factor con bloques:} & SC(\text{total}) = SC(\text{intra}) + SC(\text{bloques}) + SC(\text{tratamientos}) \end{array}$$

Generalmente no estaremos interesados en contrastar una hipótesis sobre los bloques, pero, no obstante, desearemos tener en consideración el efecto que tienen los bloques en la variable de respuesta. Al refinar el ANOVA de un factor calculando $SC(\text{bloques})$ se consigue eso y además, si los bloques se escogen sabiamente, se pueden conseguir contrastes más potentes.

El cálculo de las sumas de cuadrados se deja en general a los computadores y raramente se realiza a mano. No obstante, las fórmulas son importantes en el sentido de que revelan matemáticamente cómo se han diseñado los bloques.

Los **cuadrados medios entre bloques** se calculan de una forma similar al cálculo de $CM(\text{entre})$ del ANOVA de un factor de la Sección 11.2. Hablando en términos generales, calculamos un tipo de varianza ponderada de las medias de los bloques en el que se ponderan con el tamaño muestral del bloque las diferencias entre las media de los bloques y la media global. Si definimos el promedio de las observaciones en el bloque j como $\bar{y}_j$ y denominamos m_j al número de observaciones en el bloque j , entonces los cuadrados medios debidos a los bloques se definen como sigue:

Cuadrados medios entre bloques

$$CM(\text{bloques}) = \frac{\sum_{j=1}^J m_j (\bar{y}_{.j} - \bar{y})^2}{J - 1}$$

Análogamente a nuestras fórmulas de la Sección 11.2, definimos $SC(\text{bloques})$ como el numerador y $gl(\text{bloques})$ como denominador de $CM(\text{bloques})$, como sigue:

Suma de cuadrados y gl entre bloques

$$SC(\text{bloques}) = \sum_{j=1}^J m_j (\bar{y}_{.j} - \bar{y})^2$$

$$gl(\text{bloques}) = J - 1$$

Como se indicó anteriormente, los bloques reducen $CM(\text{intra})$. Para obtener $CM(\text{intra})$ en el experimento de bloques completos aleatorizados, se calcula

$$SC(\text{intra}) = SC(\text{total}) - SC(\text{tratamiento}) - SC(\text{bloques})$$

donde $SC(\text{tratamiento})$ y $SC(\text{total})$ se calculan como en la Sección 11.2. Como las sumas de cuadrados dan siempre valores no negativos, la fórmula anterior muestra directamente cómo los bloques reducen la variabilidad dentro de los grupos.

De forma similar, para calcular $gl(\text{intra})$ en el experimento de bloques completos aleatorizados, tenemos

$$gl(\text{intra}) = gl(\text{total}) - gl(\text{tratamiento}) - gl(\text{bloques})$$

$$= (n - 1) - (I - 1) - (J - 1)$$

$$= n - I - J + 1$$

Ejemplo 11.6.8 Alfalfa y lluvia ácida

Para los datos de crecimiento de alfalfa de la Tabla 11.6.2, el total de todas las observaciones es $1,30 + 1,15 + \dots + 0,80 = 19,04$ y la media global es

$$\bar{y} = \frac{19,04}{15} = 1,269$$

Calculamos

$$SC(\text{tratamientos}) = 5(0,910 - 1,269)^2 + 5(1,130 - 1,269)^2 + 5(1,768 - 1,269)^2 = 1,986$$

Como $I = 3$, tenemos

$$gl(\text{tratamientos}) = 3 - 1 = 2$$

de forma que

$$CM(\text{tratamientos}) = \frac{1,986}{2} = 0,993$$

Calculamos

$$SC(\text{bloques}) = 3(1,917 - 1,269)^2 + 3(1,550 - 1,269)^2$$

$$+ 3(1,077 - 1,269)^2 + 3(1,837 - 1,269)^2$$

$$+ 3(1,967 - 1,269)^2$$

$$= 2,441$$

Como $J = 5$, tenemos

$$gl(\text{bloques}) = 5 - 1 = 4$$

y

$$CM(\text{bloques}) = \frac{2,441}{4} = 0,610$$

La suma de cuadrados total se calcula como $(1,30 - 1,269)^2 + \dots + (0,80 - 1,269)^2 = 5,879$.

Restando, calculamos $SC(\text{intra})$:

$$\begin{aligned} SC(\text{intra}) &= SC(\text{total}) - SC(\text{tratamiento}) - SS(\text{bloques}) \\ &= 5,879 - 1,986 - 2,441 = 1,452 \end{aligned}$$

De forma similar, calculamos $gl(\text{intra})$ como

$$gl(\text{intra}) = gl(\text{total}) + gl(\text{tratamiento}) + gl(\text{bloques})$$

que en este caso resulta $14 - 2 - 4 = 8$.

$$\text{Por tanto, } CM(\text{intra}) = \frac{1,452}{8} = 0,182. \quad \blacksquare$$

Las sumas de cuadrados, los grados de libertad y los cuadrados medios se recogen en la tabla del ANOVA ampliada que incluye una línea para el efecto de los bloques.

Para contrastar la hipótesis nula, calculamos

$$F_s = \frac{CM(\text{tratamiento})}{CM(\text{intra})}$$

y rechazamos H_0 si el P valor es muy pequeño.

Ejemplo 11.6.9 Alfalfa y lluvia ácida

Para los datos de crecimiento de alfalfa del Ejemplo 11.6.1, el resumen del ANOVA se muestra en la Tabla 11.6.4. El estadístico F es $0,993/0,182 = 5,47$, con 2 grados de libertad para el numerador y 8 grados de libertad para el denominador. Utilizando la Tabla 10 podemos acotar el P valor como $0,02 < P \text{ valor} < 0,05$. (Utilizando un computador se obtiene un P valor de 0,0318). El P valor es pequeño, lo que indica que las diferencias entre las tres medias muestrales son mayores que las que cabría esperar solo por azar. Hay evidencia significativa a favor de que el ácido afecta al crecimiento de las plantas de alfalfa. (Merece la pena advertir que si ignoramos los bloques y realizamos erróneamente un ANOVA de un factor, obtendríamos $P \text{ valor} = 0,0842$, lo que no proporcionaría evidencia significativa a favor de que el ácido afecta, con $\alpha = 0,05$). $\blacksquare$

Tabla 11.6.4 Tabla del ANOVA para el experimento de la alfalfa

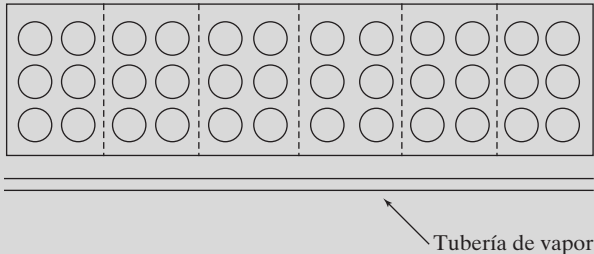
Fuente	gl	SC	CM	Razón F
Entre tratamientos	2	1,986	0,993	5,47
Entre bloques	4	2,441	0,610	
Dentro de los grupos	8	1,452	0,182	
Total	14	4,278		

Ejercicios 11.6.1-11.6.10

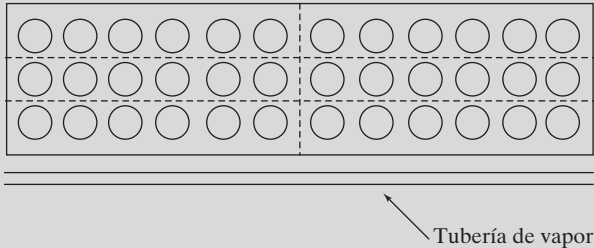
(Nota: En varios ejercicios se pide preparar una asignación aleatorizada. Para ello se puede utilizar la Tabla 1, dígitos aleatorios de su calculadora o un computador.)

11.6.1 En un experimento para comparar seis fertilizantes de tomates diferentes, se van a utilizar 36 tiestos en semillero, 6 para recibir cada fertilizante. Las plantas de tomate crecerán en invernadero y se observará la cosecha total de tomates de cada planta. El investigador ha decidido utilizar un diseño de bloques aleatorizados: los tiestos se dispondrán en seis bloques de seis plantas cada uno sobre el banco del invernadero. La figura siguiente muestra dos posibles disposiciones de los bloques.

Disposición I:



Disposición II:



Un factor que afecta a la cosecha de tomates es la temperatura, que no es exactamente constante en el invernadero. De hecho, es probable que exista un gradiente de temperatura en el banco. Se suministra calor al invernadero mediante una tubería de vapor que está situada longitudinalmente paralela a un extremo del banco, por lo que es más probable que el lado del banco más cercano a la tubería esté más caliente.

- ¿Qué disposición de bloques (I o II) es mejor? ¿Por qué?
- Prepare una asignación aleatorizada de los tratamientos a los tiestos dentro de cada bloque. (Utilice el Ejemplo 11.6.4 como guía, y suponga que las asignaciones de plantas a tiestos y de tiestos a posiciones en el bloque ya se ha realizado).

11.6.2 Se va a realizar un experimento sobre suplementos vitamínicos a lechones, utilizando las camadas como bloques en un diseño de bloques aleatorizado. Habrá cinco tra-

tamientos: cuatro tipos de suplemento y un control. Por tanto, se utilizarán cinco lechones de cada camada. El experimento incluirá cinco camadas. Prepare una asignación de bloques aleatorizados de lechones a tratamientos. (Utilice el Ejemplo 11.6.4 como guía).

11.6.3 En referencia al experimento de las vitaminas del Ejercicio 11.6.2, suponga que un colega del investigador propone un diseño alternativo: todos los cerdos de una camada dada recibirán el mismo tratamiento, asignando aleatoriamente las cinco camadas a los cinco tratamientos. Afirma que su propuesta ahorraría trabajo y simplificaría grandemente la toma de datos. ¿Si fuera el investigador, cómo contestaría a la propuesta?

11.6.4 En un estudio farmacológico sobre el comportamiento de las ratas cuando se alimentan, se asignaron aleatoriamente 18 ratas a tres grupos de tratamiento, T_1 , T_2 y T_3 . Durante la observación, los animales se mantuvieron en jaulas individuales en un estante. El estante tenía tres hileras con seis jaulas por hilera. A pesar de los esfuerzos por mantener la iluminación uniforme, las condiciones de iluminación eran ligeramente diferentes de una hilera a otra (la hilera inferior era la más oscura), y el investigador era consciente de este hecho porque se piensa que la luz influye en el comportamiento alimentario de las ratas. Se proponen los tres planes siguientes para asignar las ratas a posiciones en el estante (lo que se hará después de asignar a las ratas a grupos de tratamiento):

Plan I. Asignar aleatoriamente las 18 ratas a las 18 posiciones del estante.

Plan II. Poner todas las ratas T_1 en la primera hilera, todas las ratas T_2 en la segunda hilera y todas las ratas T_3 en la tercera hilera.

Plan III. En cada hilera, poner dos ratas T_1 , dos ratas T_2 y dos ratas T_3 .

Ordene de mejor a peor los tres planes. Explique su razonamiento.

11.6.5 Una investigadora está planeando un experimento agrícola de campo para comparar las cosechas de 25 variedades de maíz. Utilizará un diseño de bloques aleatorizados con seis bloques. Por tanto, habrá 150 parcelas, y se medirá la cosecha de cada parcela. La investigadora advierte que el tiempo requerido para cosechar todas las parcelas es tan grande que la lluvia puede interrumpir la operación. Si aparece la lluvia, puede haber una diferencia entre las cosechas antes y después de dicha lluvia. La investigadora considera los siguientes planes:

Plan I. Recoger primero todas las parcelas de la variedad 1, después todas las de la variedad 2, y así sucesivamente.

Plan II Recoger primero todas las parcelas del bloque I, después todas las del bloque II, y así sucesivamente.

¿Qué plan es mejor? ¿Por qué?

11.6.6 Para un experimento para comparar dos métodos de inseminación artificial en el ganado, están disponibles las siguientes vacas:

Novillas (14-15 meses de edad): 8 animales.

Vacas jóvenes (2-3 años de edad): 8 animales.

Vacas maduras (4-8 años de edad): 10 animales.

Los animales se van a asignar aleatoriamente a los dos grupos de tratamiento, utilizando los tres grupos de edad como bloques. Prepare una asignación adecuada, dividiendo aleatoriamente cada estrato en dos grupos iguales.

11.6.7 Verdadero o falso (e indique por qué): La razón principal para utilizar un diseño de bloques aleatorizados en un experimento es reducir el sesgo.

11.6.8 En un experimento para entender el impacto del pastoreo de peces sobre las poblaciones de invertebrados de los arroyos, unos investigadores establecieron nueve canales de observación en tres arroyos (tres canales por arroyo). Cada uno de los tres canales de un arroyo recibió tres tratamientos: no se añadieron peces, se añadieron peces Galaxias o se añadieron truchas. (Los canales incorporaban redes para evitar la entrada y la salida de peces). Doce días después de instalar los canales, se contabilizó el número de ninfas de mosca de mayo presentes en una región específica del centro del canal. El número de ninfas de cada tratamiento en cada arroyo se presenta en la tabla siguiente¹⁵:

		Arroyo		
		A	B	C
Tratamiento	Sin peces	11	8	7
	Galaxias	9	4	4
	Truchas	6	4	0

- (a) Identifique los bloques, los tratamientos (es decir, la variable explicativa de interés), y las variables de respuesta de este estudio.
- (b) En el contexto de este problema, explique a alguien que nunca haya seguido un curso de estadística cómo pueden ayudar los bloques a identificar mejor las diferencias entre los tratamientos, si existen.

11.6.9 (Continuación del Ejercicio 11.6.8)

- (a) La tabla que acompaña a este apartado es una tabla de ANOVA (no apropiada) para los datos del Ejercicio 11.6.8. Este análisis no tiene en cuenta los bloques utilizados en el experimento. Basándose en este análisis,

¿existe evidencia de que los peces afectan al número de ninfas de mosca de mayo presentes en los canales? Utilice $\alpha = 0,05$.

	gl	Suma cuadrados	Cuadrados medios	Valor F
Entre grupos	2	42,889	21,444	2,924
Dentro de los grupos	6	44,000	7,333	
Total	8	86,889		

- (b) Se presenta a continuación la tabla del ANOVA apropiada para los datos, que tiene en cuenta los bloques. Basándose en este análisis apropiado, ¿existe evidencia de que los peces afectan al número de ninfas de mosca de mayo presentes en los canales? Utilice $\alpha = 0,05$.

	gl	Suma cuadrados	Cuadrados medios	Valor F
Entre grupos	2	42,889	21,444	16,783
Entre bloques	2	38,889	19,444	15,217
Dentro de los grupos	4	5,111	1,278	
Total	8	86,889		

- (c) Calcule y compare s_{conjunta} utilizando las tablas del ANOVA de los apartados (a) y (b). ¿Por qué es una estimación mayor que la otra? ¿Qué mide s_{conjunta} en el apartado (a)? ¿Y en el apartado (b)?

11.6.10 Considere el experimento descrito en el Ejercicio 11.6.8. Además de medir el número de ninfas de mosca de mayo transcurridos los 12 días, se recogieron piedras del mismo tamaño de cada canal y se midió la masa en seco de algas (mg/cm^2) en cada una de las nueve piedras. Los resultados fueron $SC(\text{bloques}) = 0,889$, $SC(\text{intra}) = 0,444$ y $SC(\text{total}) = 2,889$.

- (a) Construya una tabla de ANOVA similar a la Tabla 11.6.4 para resumir estos datos.
- (b) ¿Existe evidencia de que la presencia o tipo de pez está asociada con la masa media en seco de algas en los canales? Utilice $\alpha = 0,05$.
- (c) ¿Puede extraerse una conclusión causal del análisis realizado en el apartado (b) basándose en estos datos? Si es así, ¿qué conclusión causal puede extraerse? Si no es así, explique por qué no es apropiado extraer una conclusión causal.

11.7 ANOVA de dos factores

ANOVA factorial

En una aplicación típica de análisis de varianza hay una única variable explicatoria o **factor** bajo estudio. Por ejemplo, en el caso de ganancia de peso del Ejemplo 11.2.1, el factor es «tipo de dieta», que puede tomar tres **niveles**: dieta 1, dieta 2 y dieta 3. Sin embargo, en algunos casos de análisis de varianza es necesario el estudio simultáneo de dos o más factores. A continuación se presenta un ejemplo.

Ejemplo 11.7.1 Crecimiento de plantas de soja

Un fisiólogo de plantas investigó el efecto del estrés mecánico en el crecimiento de plantas de soja. Se asignaron aleatoriamente plantas en tiestos individuales a cuatro grupos de tratamiento de 13 plantas cada uno. Las plantas de dos grupos fueron sometidas a estrés agitándolas durante 20 minutos dos veces al día, mientras que los dos grupos de control no se sometieron a estrés. Por tanto, el primer factor en el experimento era la presencia o ausencia de estrés, con dos niveles: control o estrés. Además, las plantas se hicieron crecer con luz baja o moderada. Por tanto, el segundo factor fue la cantidad de luz, con dos niveles: luz baja o luz moderada. Este experimento es un ejemplo de *experimento factorial de 2×2* . Incluye cuatro tratamientos:

- Tratamiento 1: control, luz baja
- Tratamiento 2: estrés, luz baja
- Tratamiento 3: control, luz moderada
- Tratamiento 4: estrés, luz moderada

Después de 16 días de crecimiento, las plantas se recogieron y se midió el área total de las hojas (cm^2) de cada planta. Los resultados se muestran en la Tabla 11.7,1 y se representan en la Figura 11.7.1¹⁶.

Tabla 11.7.1 Área de las hojas (cm^2) de plantas de soja				
	Tratamiento			
	Control, luz baja	Estrés, luz baja	Control, luz moderada	Estrés, luz moderada
	264	235	314	283
	200	188	320	312
	225	195	310	291
	268	205	340	259
	215	212	299	216
	241	214	268	201
	232	182	345	267
	256	215	271	326
	229	272	285	241
	288	163	309	291
	253	230	337	269
	288	255	282	282
	230	202	273	257
Media	245,3	212,9	304,1	268,8
DT	27,0	29,7	26,9	35,2
n	13	13	13	13

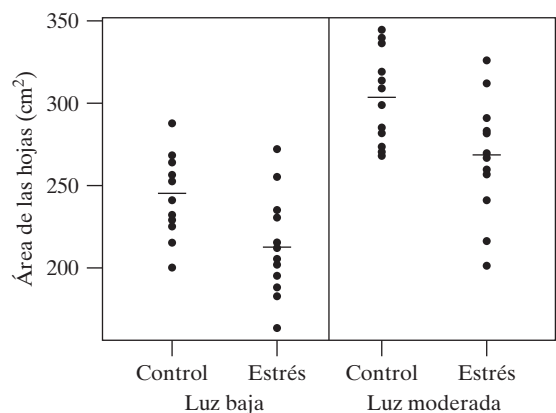


Figura 11.7.1 Área de las hojas de plantas de soja que han recibido cuatro tratamientos diferentes. Las medias de los grupos se indican como (—)

La Figura 11.7.1 muestra evidencia de que el estrés reduce el área de las hojas. Esto es cierto bajo condiciones de luz baja y de luz moderada. Asimismo, la luz moderada incrementa el área de las hojas, tanto si hay como si no hay estrés. ■

Un modelo para esta situación es

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

donde y_{ijk} es la k -ésima observación del nivel i del primer factor y del nivel j del segundo factor. El término τ_i representa el efecto del nivel i del primer factor (condición de estrés en el Ejemplo 11.7.1) y el término β_j representa ahora el efecto del nivel j del segundo factor (condición de iluminación en el Ejemplo 11.7.1).

Cuando se estudian dos factores dentro de un único experimento resulta de ayuda organizar las medias muestrales en una tabla que refleje la estructura del experimento y presentar las medias en una gráfica que ponga de relieve dicha estructura.

Ejemplo 11.7.2 Crecimiento de plantas de soja

La Tabla 11.7.2 resume los datos del Ejemplo 11.7.1. Por ejemplo, cuando el primer factor está en su primer nivel (control) y el segundo factor está en su primer nivel (luz baja) la media muestral es $\bar{y}_{11} = 245,3$. El formato de esta tabla nos permite considerar fácilmente los dos factores (condición de estrés y condición de iluminación) de forma separada y conjunta. La última columna muestra el efecto de la luz para cada nivel de estrés. Los números de esta columna confirman la impresión visual de la Figura 11.7.1. La luz moderada aumenta el área media de las hojas aproximadamente en la misma cantidad cuando las plantas se someten a estrés y cuando no se someten a estrés. Asimismo, la última fila (−32,4 frente a −35,3) muestra que el efecto del estrés es aproximadamente el mismo en cada nivel de iluminación. ■

Tabla 11.7.2 Medias de las hojas en el experimento de las plantas de soja				
		Condición de iluminación		Diferencia
		Luz baja	Luz moderada	
Condición de agitación	Control	245,3	304,1	58,8
	Estrés	212,9	268,8	55,9
	Diferencia	−32,4	−35,3	

Si la influencia conjunta de dos factores es igual a la suma de sus influencias separadas, se dice que los dos factores son **aditivos** en sus efectos. Por ejemplo, consideremos el experimento de las plantas de soja del Ejemplo 11.7.1. Si el estrés reduce el área media de las hojas en la misma cantidad para cualquier condición de iluminación, entonces el efecto del estrés (un efecto negativo en este caso) se *suma* al efecto de la luz. Para visualizar esta aditividad de efectos, consideremos la Figura 11.7.2, que muestra los datos con las cuatro medias de los tratamientos. Las líneas continuas que unen las medias de los tratamientos son casi paralelas porque los datos presentan un patrón de aditividad casi perfecta*.

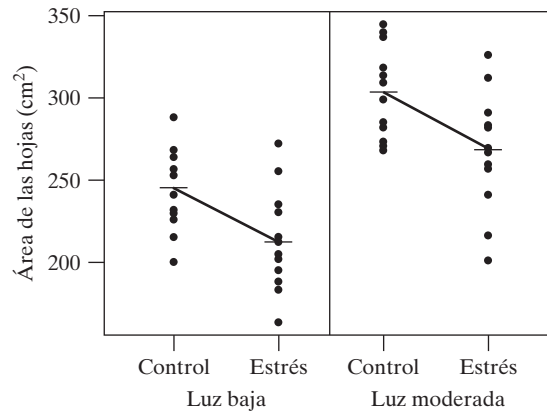


Figura 11.7.2 Datos y medias de tratamientos para el experimento de las plantas de soja

Cuando los efectos de los factores son aditivos se dice que no hay **interacción** entre los factores. Una gráfica que muestra solo las medias de los tratamientos se denomina a menudo gráfica de interacción. La Figura 11.7.3, que es una versión resumida de la Figura 11.7.2, es una gráfica de interacción que resalta el efecto del estrés en el área media de las hojas para las dos condiciones de iluminación. Se pueden hacer gráficas análogas para desplazar el foco de atención a la comparación del efecto de la luz en el área media de las hojas para las dos condiciones de estrés.

Algunas veces el efecto que tiene un factor sobre una variable de respuesta depende del nivel de un segundo factor. Cuando esto sucede se dice que los dos factores interactúan en su efecto sobre la respuesta. A continuación se presenta un ejemplo.

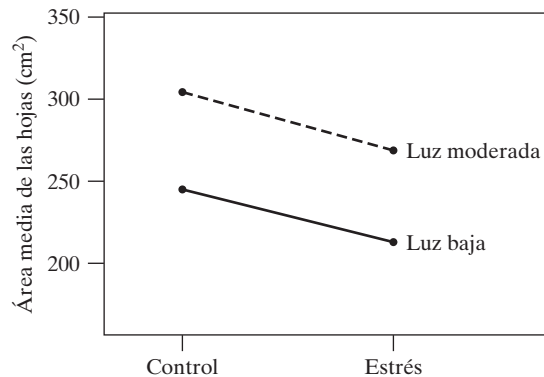


Figura 11.7.3 Gráfica de interacción para el experimento de las plantas de soja

* La diferencia entre el área media de las hojas en el caso de estrés con la condición de luz baja (212,9) y el área media de las hojas para el caso de control con la condición de luz baja (245,3) se denomina **efecto simple** de la agitación bajo condiciones de luz baja. Por tanto, el efecto simple de la agitación en condiciones de luz baja es $212,9 - 245,3 = -32,4$. Asimismo, el efecto simple de la agitación bajo condiciones de luz moderada es $268,8 - 304,1 = -35,3$. Un **efecto principal** es un promedio de efectos simples. Por ejemplo, el efecto principal de la agitación es $(-32,4 + -35,3)/2 = -33,85$. El efecto principal de la luz es $(58,8 + 55,9)/2 = 57,35$.

Ejemplo 11.7.3 Suplementos de hierro en bebidas de fruta con leche

La complementación con hierro y cinc de bebidas de fruta con leche es una práctica común. Para entender mejor los efectos de la complementación de las bebidas en la retención celular del hierro, unos investigadores realizaron un experimento complementando bebidas de fruta con leche con niveles bajos y altos de hierro (Fe) y cinc (Zn). Las bebidas se digirieron en un tracto gastrointestinal simulado y se midió la retención celular de hierro ($\mu\text{g Fe/mg}$ de proteína celular). La Tabla 11.7.3 resume los datos, que incluían ocho observaciones para cada combinación de niveles de suplemento de Fe y de Zn¹⁷. La Figura 11.7.4 es un gráfico de interacción que muestra las cuatro medias. Nótese que cuando el nivel de suplemento de Zn es bajo, el efecto del suplemento de Fe en la retención celular es mucho menor que cuando el nivel de suplemento de Zn es alto (es decir, las pendientes de las dos líneas son diferentes, las líneas no son paralelas). Por tanto, el efecto del suplemento de Fe en la retención celular media depende de la cantidad de suplemento de Zn utilizado. Se dice que el Fe y el Zn interactúan en sus efectos sobre la retención celular. ■

		Nivel de Zn		Diferencia
		Bajo	Alto	
Nivel de Fe	Bajo	0,707	0,215	− 0,492
	Alto	0,994	1,412	0,418
	Diferencia	0,287	1,197	

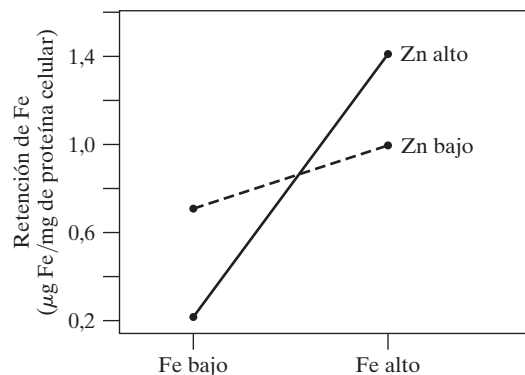


Figura 11.7.4 Gráfica de interacción para el experimento de suplemento de bebidas

Cuando se sospecha que dos factores interactúan en un escenario de ANOVA, se puede ampliar nuestro modelo añadiendo un término de interacción:

$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

Aquí, el término γ_{ij} es el efecto de interacción entre el nivel i del primer factor y el nivel j del segundo factor. Como antes, si hay n observaciones en total, entonces $\text{gl}(\text{total}) = n - 1$. Si hay I niveles del primer factor, entonces tiene $I - 1$ grados de libertad. Asimismo, si hay J niveles de segundo factor, entonces tiene $J - 1$ grados de libertad. Hay $(I - 1) \times (J - 1)$ grados de libertad de interacción. Con I niveles del primer factor y J niveles de segundo factor hay IJ combinaciones de tratamientos. Por tanto, $\text{gl}(\text{intra}) = n - IJ^*$.

* Esto es análogo a la definición de $\text{gl}(\text{intra}) = n - 1$ para el ANOVA de un factor de la Sección 11.2. En cada escenario, $\text{gl}(\text{intra}) = \text{número total de observaciones} - \text{número de tratamientos}$.

Una hipótesis nula de interés es que todos los términos de interacción son cero:

$$H_0: \gamma_{11} = \gamma_{12} = \dots = \gamma_{IJ} = 0$$

Para contrastar este hipótesis nula calculamos

$$F_s = \frac{\text{CM(interacción)}}{\text{CM(intra)}}$$

y rechazamos H_0 si el P valor es muy pequeño.

Ejemplo 11.7.4 Suplementos de hierro en bebidas de fruta con leche

La Tabla 11.7.4 muestra los resultados del análisis de varianza para el experimento de suplemento en las bebidas del Ejemplo 11.7.3. Esta tabla incluye una línea para el término de interacción*. Había ocho observaciones en cada combinación de niveles de suplemento de Fe y Zn. Por tanto, $n = 32$ y $\text{gl}(\text{total}) = 31$. En este ejemplo, $I = J = 2$, por lo que $\text{gl}(\text{niveles de Fe}) = \text{gl}(\text{niveles de Zn}) = \text{gl}(\text{interacción}) = 1$. Podemos calcular $\text{gl}(\text{intra})$ restando: $\text{gl}(\text{intra}) = 31 - 1 - 1 - 1 = 28$ (esto coincide con la fórmula de $\text{gl}(\text{intra}) = n - IJ = 32 - 2 \times 2$).

Para comprobar si los niveles de suplemento de Fe y de Zn interactúan, utilizamos la razón F que vale $1,6555/0,0019 = 871,3$, que tienen un grado de libertad 1 para el numerador y 28 para el denominador. Utilizando la Tabla 10 podemos acotar el P valor como $P \text{ valor} < 0,0001$. El P valor es extremadamente pequeño, lo que indica que el patrón de interacción que se ve en la Figura 11.7.4 es más pronunciado del que cabría esperar por el mero azar. Por tanto, rechazamos H_0 . ■

Tabla 11.7.4 Tabla de ANOVA para el experimento de suplemento de bebidas				
Fuente	gl	SC	CM	Razón F
Entre niveles de Fe	1	4,4023	4,4023	2,317,0
Entre niveles de Zn	1	0,0109	0,0109	5,736
Interacción	1	1,6555	1,6555	871,3
Dentro de los grupos	28	0,0523	0,0019	
Total	31	6,1210		

El concepto de interacción aparece en biología. Los términos «sinergia» y «antagonismo» describen interacciones entre agentes biológicos. El término «epístasis» describe la interacción entre genes en dos loci.

Cuando hay interacciones, como en el Ejemplo 11.7.3, los efectos principales de los factores no tienen sus interpretaciones habituales. Considerando el Ejemplo 11.7.3, es difícil plantear el efecto independiente del Fe porque la naturaleza y magnitud de dicho efecto depende del nivel particular de suplemento de Zn. Por ello, en general se comprueba primero la presencia de interacciones. Si existen, como en el ejemplo de suplemento de bebidas, entonces a menudo el análisis se detiene en este punto. Si no se encuentra evidencia de la existencia de interacciones (es decir, si no se rechaza H_0), entonces procederemos a contrastar los efectos principales de los factores individuales. El ejemplo siguiente ilustra este proceso.

Ejemplo 11.7.5 Crecimiento de plantas de soja

La Tabla 11.7.5 es una tabla de análisis de la varianza para los datos de crecimiento de plantas de soja del Ejemplo 11.7.1. La hipótesis nula

$$H_0: \gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22} = 0$$

* Las fórmulas de ANOVA que se utilizan para calcular la suma de cuadrados debida a la interacción son más bien complicadas y no se presentan aquí. En particular, es importante si el diseño es o no «equilibrado». El experimento de suplemento en las bebidas es equilibrado en el sentido de que hay ocho observaciones en cada una de las cuatro combinaciones de niveles de los factores que se muestran en la Tabla 11.7.3. Sin embargo, son posibles diseños no equilibrados, que conducen a cálculos y análisis complicados. Aquí nos basamos en el *software* estadístico para calcular las sumas de cuadrados necesarias.

se contrasta con la razón F

$$F_s = \frac{\text{CM(interacción)}}{\text{CM(intra)}} = \frac{26,3}{895,34} = 0,029$$

Buscando en la Tabla 10 con 1 y 12 grados de libertad, vemos que el P valor es mayor que 0,20. Por tanto, no hay evidencia significativa de interacción y no rechazamos H_0 .

Como no hay evidencia de interacciones, contrastamos el efecto principal del nivel de estrés. En este caso la razón F es

$$F_s = \frac{\text{CM(entre niveles de estrés)}}{\text{CM(intra)}} = \frac{14.858,5}{895,34} = 16,6$$

Esto es altamente significativo (es decir, el P valor es muy pequeño) y rechazamos H_0 .

Asimismo, el contraste para el efecto principal de los niveles de iluminación tiene una razón F de

$$F_s = \frac{\text{CM(entre niveles de iluminación)}}{\text{CM(intra)}} = \frac{42.751,6}{895,34} = 47,75$$

De nuevo, esto es altamente significativo, y rechazamos H_0 . ■

Tabla 11.7.5 Tabla de ANOVA para el experimento de crecimiento de plantas de soja				
Fuente	gl	SC	CM	Razón F
Entre niveles de estrés	1	14.858,5	14.858,5	16,60
Entre niveles de iluminación	1	42.751,6	42.751,6	47,75
Interacción	1	26,3	26,3	0,029
Dentro de los grupos	48	42.976,3	895,34	
Total	51	100.612,7		

Cuando hay más de dos niveles de un factor se pueden utilizar gráficas de interacción, como en el ejemplo siguiente.

Ejemplo 11.7.6 Sapos

Unos investigadores estudiaron el efecto que la exposición a la radiación ultravioleta-B tenía sobre la supervivencia de embriones del sapo del oeste *Bufo boreas*. Para ello, realizaron un experimento en el que varios embriones de *Bufo boreas* se situaron a tres niveles de profundidad (10 cm, 50 cm o 100 cm) y se sometieron a dos niveles de radiación (expuestos a UV-B o protegidos). La variable de respuesta fue el porcentaje de embriones que sobrevivieron a la incubación. La Tabla 11.7.6 resume los datos, que incluían cuatro observaciones para cada combinación de profundidad y exposición a UV-B. La Figura 11.7.5 es una gráfica de interacción que muestra las seis medias. La presencia de interacciones se hace aparente en la gráfica. La Tabla 11.7.7 resume el análisis de varianza¹⁸. ■

Tabla 11.7.6 Porcentaje de embriones supervivientes en el experimento de los sapos				
		UV-B		
		Expuestos	Protegidos	Diferencia
Profundidad en el agua	10 cm	0,425	0,759	0,334
	50 cm	0,729	0,748	0,019
	100 cm	0,785	0,766	−0,019

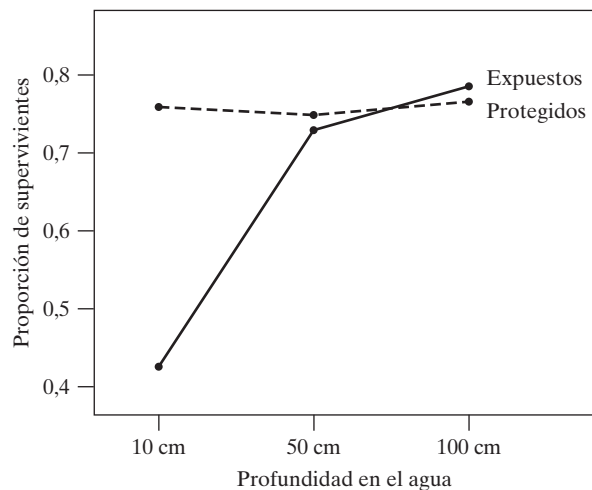


Figura 11.7.5 Gráfica de interacción para el experimento de los sapos

Tabla 11.7.7 Tabla de ANOVA para el experimento de los sapos				
Fuente	gl	SC	CM	Razón <i>F</i>
Entre profundidades	2	0,150676	0,075338	13,92
Entre niveles de UV-B	1	0,074371	0,074371	13,74
Interacción	2	0,150185	0,075093	13,88
Dentro de los grupos	18	0,097401	0,005411	
Total	23	0,472633		

El tema de las interacciones se considera también en la Sección 11.8.

Ejercicios 11.7.1-11.7.6

11.7.1 Un fisiólogo de plantas investigó el efecto de la inundación en el metabolismo de dos especies de árbol: el abedul de río, tolerante a la inundación, y el abedul europeo, no tolerante. Se inundaron durante un día cuatro plantas de cada especie, mientras que otras cuatro se utilizaron como control. Se midió la concentración de adenosina trifosfato (ATP) en las raíces de cada planta. Los datos (nmol ATP por mg de tejido) se muestran en la tabla siguiente¹⁹.

	Abedul de río		Abedul europeo	
	Inundación	Control	Inundación	Control
	1,45	1,70	0,21	1,34
	1,19	2,04	0,58	0,99
	1,05	1,49	0,11	1,17
	1,07	1,91	0,27	1,30
Media	1,19	1,785	0,2925	1,20

Prepare una gráfica de interacción (como la Figura 11.7.3).

11.7.2 Considere los datos del Ejercicio 11.7.1. Para esos datos, $SC(\text{especies de abedul}) = 2,19781$, $SC(\text{inundación}) = 2,25751$, $SC(\text{interacción}) = 0,097656$ y $SC(\text{intra}) = 0,47438$.

- Construya la tabla del ANOVA.
- Realice un test *F* para las interacciones. Utilice $\alpha = 0,05$.
- Contraste la hipótesis nula de que la especie no tiene efecto en la concentración de ATP. Utilice $\alpha = 0,01$.
- Suponiendo que las cuatro poblaciones tienen la misma desviación típica, utilice los datos para obtener una estimación de dicha desviación típica.

11.7.3 Se realizó una prueba clínica completamente aleatorizada y doble ciego para comparar la efectividad de dos medicamentos, ticrinafeno (T) e hidroclorotiazida (H), en el tratamiento de la presión sanguínea alta. Cada medicamento se administró a dosis baja o dosis alta durante seis semanas. La tabla que acompaña al ejercicio muestra los resultados de descenso de la presión sanguínea sistólica (punto de partida menos valor final)²⁰.

	Ticrinafeno (T)		Hidroclorotiazida (H)	
	Dosis baja	Dosis alta	Dosis baja	Dosis alta
Media	13,9	17,1	15,8	17,5
Número de pacientes	53	57	55	58

Prepare un gráfico de interacción (como la Figura 11.7.3).

11.7.4 Considere los datos del Ejercicio 11.7.3. La diferencia de respuesta entre T y H parece ser mayor a dosis bajas que a dosis altas.

- Realice un test F de interacciones para evaluar si este patrón puede ser atribuido a la variación por azar. Utilice $\alpha = 0,10$. Para estos datos, $SC(\text{interacción}) = 31,33$ y $SC(\text{intra}) = 30.648,81$.
- Basándose en sus resultados del apartado (a), ¿es sensato examinar e interpretar los efectos principales del medicamento y de la dosis?

11.7.5 Considere los datos del Ejercicio 11.7.3. Para estos datos, $SC(\text{medicamento}) = 69,22$, $SC(\text{dosis}) = 330,00$, $SC(\text{interacción}) = 31,33$ y $SC(\text{intra}) = 30.648,81$.

- Construya la tabla del ANOVA.
- Realice un contraste de la hipótesis nula de que los efectos de los dos medicamentos (T y H) son iguales. Utilice $\alpha = 0,05$.

11.7.6 En un estudio sobre crecimiento de lechugas, se asignaron aleatoriamente 36 plantas de semillero para recibir iluminación alta o iluminación baja y para crecer en una solución de nutrientes estándar o en una que contenía nitrógeno extra. Tras 16 días de crecimiento, se recogieron las plantas y se obtuvo el peso en seco de las hojas de cada planta. La tabla que acompaña al ejercicio muestra el peso medio en seco de las hojas (g) de las 9 plantas de cada grupo de tratamiento²¹.

	Solución de nutrientes	
	Estándar	Nitrógeno extra
Iluminación baja	2,16	3,09
Iluminación alta	3,26	4,48

Para estos datos, $SC(\text{solución de nutrientes}) = 10,4006$, $SC(\text{iluminación}) = 13,95023$, $SC(\text{interacción}) = 0,18923$ y $SC(\text{intra}) = 11,1392$.

- Construya la tabla del ANOVA.
- Realice un test F de interacciones. Utilice $\alpha = 0,05$.
- Contraste la hipótesis nula de que la solución de nutrientes no tiene efecto en el peso. Utilice $\alpha = 0,01$.

11.8 Combinaciones lineales de medias (opcional)

En muchos estudios se pueden plantear cuestiones interesantes considerando combinaciones lineales de medias de grupos. Una **combinación lineal** L es una cantidad con la forma

$$L = m_1\bar{y}_1 + m_2\bar{y}_2 + \cdots + m_I\bar{y}_I$$

siendo las m los coeficientes de las $\bar{y}_i$.

Combinaciones lineales para ajuste

Un uso de las combinaciones lineales es «ajustar» una variable externa, como ilustra el siguiente ejemplo.

Ejemplo 11.8.1 Capacidad vital forzada (CVF)

Una medida de la función pulmonar es la capacidad vital forzada (CVF), que es la máxima cantidad de aire que una persona puede expirar en una respiración. En un estudio sobre la salud pública, unos investigadores midieron la CVF de una muestra grande de personas. Los resultados para varones ex fumadores, estratificados por edad, se muestran en la Tabla 11.8.1²².

Supongamos que se desea calcular un valor resumen de la CVF en varones ex fumadores. Una posibilidad sería simplemente calcular la media global de los 481 valores observados, que es 4,56 litros. Pero la media global tiene un serio inconveniente: no se puede comparar de modo significativo con otras poblaciones que tengan diferentes distribuciones de edad. Por ejemplo, supongamos que deseamos comparar ex fumadores con no fumadores. La diferencia

Tabla 11.8.1 CVF en varones ex fumadores			
CVF (litros)			
Edad (años)	<i>n</i>	Media	DT
25-34	83	5,29	0,76
35-44	102	5,05	0,77
45-54	126	4,51	0,74
55-64	97	4,24	0,80
65-74	73	3,58	0,82
25-74	481	4,56	

en los valores de la CVF estaría distorsionada porque los ex fumadores son un grupo con mayor edad (lo cual no es sorprendente) que los no fumadores. Una medida resumen que no tiene esta desventaja es la media «ajustada por edad», que es un estimador de la CVF media en una población de referencia con una distribución de edad específica. Para ilustrar esta idea, usaremos la distribución de referencia de la Tabla 11.8.2, que es (aproximadamente) la distribución de la población completa de Estados Unidos²³.

Tabla 11.8.2 Distribución de la edad en la población de referencia	
Edad	Frecuencia relativa
25-34	0,23
35-44	0,22
45-54	0,24
55-64	0,22
65-74	0,09

La media «ajustada por edad» es la siguiente combinación lineal:

$$L = 0,23\bar{y}_1 + 0,22\bar{y}_2 + 0,24\bar{y}_3 + 0,22\bar{y}_4 + 0,09\bar{y}_5$$

Nótese que los coeficientes (*m*) son las frecuencias relativas de la población de referencia. Utilizando la Tabla 11.8.1, el valor de *L* es

$$L = (0,23)(5,29) + (0,22)(5,05) + (0,24)(4,51) + (0,22)(4,24) + (0,09)(3,58) = 4,67 \text{ litros}$$

Este valor es una estimación de la CVF media en una población idealizada de personas que son biológicamente como los varones ex fumadores, pero con la distribución de edad de la población de referencia. ■

Contrastes lineales

Una combinación lineal cuyos coeficientes (*m*) suman cero se denomina **contraste**. El ejemplo siguiente muestra cómo se pueden utilizar los contrastes para describir los resultados de un experimento.

Ejemplo 11.8.2 Crecimiento de plantas de soja

La Tabla 11.8.3 muestra las medias de los tratamientos y los tamaños muestrales del experimento de crecimiento de plantas de soja del Ejemplo 11.6.8. Podemos utilizar contrastes para describir los efectos del estrés bajo las dos condiciones de iluminación.

Tabla 11.8.3 Datos de crecimiento de plantas de soja

Tratamiento	Área media de las hojas (cm ²)	n
1. Control, iluminación baja	245,3	13
2. Estrés, iluminación baja	212,9	13
3. Control, iluminación moderada	304,1	13
4. Estrés, iluminación moderada	268,8	13

- (a) Nótese en primer lugar que una diferencia ordinaria de parejas es un contraste. Por ejemplo, para medir el efecto del estrés con iluminación baja podemos considerar el contraste

$$L = \bar{y}_1 - \bar{y}_2 = 245,3 - 212,9 = 32,4$$

Para este contraste, los coeficientes son $m_1 = 1$, $m_2 = -1$, $m_3 = 0$ y $m_4 = 0$. Nótese que su suma es cero.

- (b) Para medir el efecto del estrés con iluminación moderada podemos considerar el contraste

$$L = \bar{y}_3 - \bar{y}_4 = 304,1 - 268,8 = 35,3$$

Para este contraste, los coeficientes son $m_1 = 0$, $m_2 = 0$, $m_3 = 1$ y $m_4 = -1$.

- (c) Para medir el efecto global del estrés, podemos promediar los contrastes de los apartados (a) y (b), obteniendo el contraste

$$\begin{aligned} L &= \frac{1}{2} (\bar{y}_1 - \bar{y}_2) + \frac{1}{2} (\bar{y}_3 - \bar{y}_4) \\ &= \frac{1}{2} (32,4) + \frac{1}{2} (35,3) = 33,85 \end{aligned}$$

Para este contraste, los coeficientes son $m_1 = \frac{1}{2}$, $m_2 = -\frac{1}{2}$, $m_3 = \frac{1}{2}$ y $m_4 = -\frac{1}{2}$. ■

Error típico de una combinación lineal

Toda combinación lineal L es un estimador, basado en los valores de $\bar{y}_i$, de la correspondiente combinación lineal de las medias poblacionales (μ). Como base para la inferencia estadística, es necesario considerar el error típico de una combinación lineal, que se calcula como sigue.

Error típico de L

El error típico de la combinación lineal

$$L = m_1 \bar{y}_1 + m_2 \bar{y}_2 + \dots + m_I \bar{y}_I$$

es

$$ET_L = s_{\text{conjunta}} \sqrt{\sum_{i=1}^I \frac{m_i^2}{n_i}}$$

siendo $s_{\text{conjunta}} = \sqrt{\text{CM}(\text{intra})}$ del ANOVA.

El ET se puede expresar explícitamente como

$$ET_L = s_{\text{conjunta}} \sqrt{\left(\frac{m_1^2}{n_1} + \frac{m_2^2}{n_2} + \dots + \frac{m_I^2}{n_I} \right)}$$

Si todos los tamaños muestrales (n_i) son iguales, el ET se puede expresar como

$$ET_L = s_{\text{conjunta}} \sqrt{\frac{(m_1^2 + m_2^2 + \dots + m_I^2)}{n}} = s_{\text{conjunta}} \sqrt{\frac{1}{n} \sum_{i=1}^I m_i^2}$$

Los dos ejemplos siguientes ilustran la aplicación de la fórmula del error típico.

Ejemplo 11.8.3 Capacidad vital forzada (CVF)

Para la combinación lineal L definida en el Ejemplo 11.8.1, obtenemos que

$$\sum_{i=1}^I \frac{m_i^2}{n_i} = \frac{0,23^2}{83} + \frac{0,22^2}{102} + \frac{0,24^2}{126} + \frac{0,22^2}{97} + \frac{0,09^2}{73} = 0,0021789$$

El ANOVA para esos datos da $s_{\text{conjunta}} = \sqrt{0,59989} = 0,77453$. Por tanto, el error típico de L es

$$ET_L = 0,77453 \sqrt{0,0021789} = 0,0362$$

Ejemplo 11.8.4 Crecimiento de plantas de soja

Para la combinación lineal L definida en el Ejemplo 11.8.2(a), obtenemos que

$$\sum_{i=1}^I m_i^2 = (1)^2 + (-1)^2 + (0)^2 + (0)^2 = 2$$

de forma que

$$ET_L = s_{\text{conjunta}} \sqrt{\frac{2}{13}}$$

Intervalos de confianza

Las combinaciones lineales de medias se pueden utilizar para contrastar hipótesis y para construir intervalos de confianza. Los valores críticos se obtienen de la distribución t de Student con

$$gl = gl(\text{intra})$$

del ANOVA*. Los intervalos de confianza se construyen utilizando el formato habitual de la t de Student. Por ejemplo, el intervalo de confianza del 95 % es

$$L \pm t_{0,025} ET_L$$

El ejemplo siguiente ilustra la construcción del intervalo de confianza.

Ejemplo 11.8.5 Crecimiento de plantas de soja

Considere el contraste definido en el Ejemplo 11.8.2(c):

$$L = \frac{1}{2} (\bar{y}_1 - \bar{y}_2) + \frac{1}{2} (\bar{y}_3 - \bar{y}_4)$$

Este contraste es un estimador del valor

$$\lambda = \frac{1}{2} (\mu_1 - \mu_2) + \frac{1}{2} (\mu_3 - \mu_4)$$

* Este método para determinar valores críticos no tiene en cuenta las comparaciones múltiples. Véase la Sección 11.9.

que se puede interpretar como el valor verdadero (poblacional) del estrés, promediado sobre todas las condiciones de iluminación. Construyamos un intervalo de confianza del 95 % para esta diferencia verdadera.

En el Ejemplo 11.8.2 obtuvimos que el valor de L es

$$L = 33,85$$

Para calcular ET_L , primero calculamos

$$\sum_{i=1}^I \frac{m_i^2}{n_i} = \frac{(\frac{1}{2})^2}{13} + \frac{(-\frac{1}{2})^2}{13} + \frac{(\frac{1}{2})^2}{13} + \frac{(-\frac{1}{2})^2}{13} = \frac{1}{13}$$

Del ANOVA, que se muestra en la Tabla 11.8.4, obtenemos que $s_{conjunta} = \sqrt{895,34} = 29,922$. Por tanto,

$$ET_L = s_{conjunta} \sqrt{\sum_{i=1}^I \frac{m_i^2}{n_i}} = 29,922 \sqrt{\frac{1}{13}} = 8,299$$

Tabla 11.8.4 Tabla de ANOVA para el experimento de crecimiento de plantas de soja				
Fuente	gl	SC	CM	Razón F
Entre niveles de estrés	1	14.858,5	14.858,5	16,60
Entre niveles de iluminación	1	42.751,6	42.751,6	47,75
Interacción	1	26,3	26,3	0,029
Dentro de los grupos	48	42.976,3	895,34	
Total	51	100.613		

Utilizando la Tabla 4 con $gl = 40 \approx 48$, obtenemos $t_{40,0,025} = 2,021$. El intervalo de confianza es

$$33,85 \pm (2,021)(8,299)$$

$$33,85 \pm 16,77$$

o (17,1, 50,6).

Tenemos una confianza del 95 % en que el efecto de estrés, promediado sobre todas las condiciones de iluminación, reduce el área de las hojas en una cantidad cuya media está entre 17,1 cm^2 y 50,6 cm^2 . ■

Tests t

Para contrastar la hipótesis nula de que el valor poblacional de un contraste es cero, el estadístico de contraste se calcula como

$$t_s = \frac{L}{ET_L}$$

y el test t se realiza de la forma habitual. El test t se ilustrará en el Ejemplo 11.8.6.

Contrastes lineales para evaluar la interacción

Algunas veces los investigadores desean estudiar los efectos separados y conjuntos de dos o más factores sobre una variable de respuesta Y . En la Sección 11.7 se presentó el concepto de interacción entre dos factores. Los contrastes lineales proporcionan otra forma de estudiar estas interacciones. Se presenta a continuación un ejemplo.

Ejemplo 11.8.6 Crecimiento de plantas de soja

En el experimento sobre crecimiento de plantas de soja (Ejemplo 11.6.8 y Ejemplo 11.8.2), los dos factores de interés son la condición de estrés y el nivel de iluminación. La Tabla 11.8.5 muestra las medias de los tratamientos, dispuestas en un nuevo formato que nos permite considerar fácilmente los factores en conjunto y por separado.

		Condición de iluminación		
		Iluminación baja	Iluminación moderada	Diferencia
Condición de agitación	Control	245,3 (1)	304,1 (3)	58,8
	Estrés	212,9 (2)	268,8 (4)	55,9
	Diferencia	- 32,4	- 35,3	

Para cada nivel de iluminación, el efecto medio del estrés se puede medir mediante un contraste:

$$\text{Efecto del estrés con iluminación baja: } \bar{y}_2 - \bar{y}_1 = 212,9 - 245,3 = -32,4$$

$$\text{Efecto del estrés con iluminación moderada: } \bar{y}_4 - \bar{y}_3 = 268,8 - 304,1 = -35,3$$

Consideremos ahora la pregunta: ¿Es la reducción del área de las hojas debida al estrés la misma bajo las dos condiciones de iluminación? Una forma de abordar esta cuestión es comparar $(\bar{y}_2 - \bar{y}_1)$ frente a $(\bar{y}_4 - \bar{y}_3)$. La diferencia de estos dos valores es un contraste lineal

$$\begin{aligned} L &= (\bar{y}_2 - \bar{y}_1) - (\bar{y}_4 - \bar{y}_3) \\ &= -32,4 - (-35,3) = 2,9 \end{aligned}$$

Este contraste L se puede utilizar como base para obtener un intervalo de confianza o para realizar un contraste de hipótesis. La hipótesis nula es

$$H_0: (\mu_2 - \mu_1) = (\mu_4 - \mu_3)$$

o, expresada con palabras,

H_0 : el efecto del estrés es el mismo bajo las dos condiciones de iluminación.

Para la combinación lineal L anterior, $\sum_{i=1}^I \frac{m_i^2}{n_i} = \frac{4}{13}$, y el error típico es

$$ET_L = s_{\text{conjunta}} \sqrt{\sum_{i=1}^I \frac{m_i^2}{n_i}} = s_{\text{conjunta}} \sqrt{\left(\frac{4}{13}\right)} = 29,922 \sqrt{\frac{4}{13}} = 16,6$$

El estadístico de contraste es

$$t_s = \frac{2,9}{16,6} = 0,2$$

Utilizando la Tabla 4 con $gl = 40$ obtenemos $t_{40,0,20} = 1,303$. Los datos no proporcionan virtualmente ninguna evidencia a favor de que el efecto del estrés es diferente bajo las dos condiciones de iluminación. Esto es consistente con el test F de interacciones realizado en el Ejemplo 11.7.5. ■

La definición estadística de interacción presentada en la Sección 11.7 y vista a través de la lente del contraste aquí es más bien especializada. Se define en términos de la variable observada en vez de en términos de un mecanismo biológico. Además, la interacción medida por un contraste se define mediante *diferencias* entre medias. En algunas aplicaciones, el biólogo podría pensar que la razón de medias tiene un sentido más relevante que las diferencias. El ejemplo siguiente muestra cómo los dos puntos de vista pueden conducir a interpretaciones diferentes.

Ejemplo 11.8.7 Aberraciones cromosómicas

Un equipo investigó los efectos separados y conjuntos en ratones de la exposición a alta temperatura (35 °C) e inyección del medicamento para el cáncer ciclofosfamida (CTX). Se utilizó un diseño completamente aleatorizado, con ocho ratones en cada grupo de tratamiento. En cada animal, los investigadores midieron la incidencia de cierta aberración cromosómica en la médula ósea. El resultado se expresa como número de células anormales por cada 1.000 células. Las medias de los tratamientos se muestran en la Tabla 11.8.6²⁴.

Tabla 11.8.6 Incidencia media de aberraciones cromosómicas siguiendo varios tratamientos			
		Inyección	
		CTX	Ninguna
Temperatura	Ambiente	23,5	2,7
	Alta	75,4	20,9

¿Es el efecto observado de la CTX mayor a temperatura ambiente o a temperatura alta? La respuesta depende de si el «efecto» se mide de forma absoluta o relativa.

Medido como diferencia, el efecto de la CTX es

$$\text{Temperatura ambiente: } 23,5 - 2,7 = 20,8$$

$$\text{Temperatura alta: } 75,4 - 20,9 = 54,5$$

Por tanto, el efecto absoluto de la CTX es mayor a alta temperatura. Sin embargo, esta relación se invierte si expresamos el efecto de la CTX como una razón en lugar de una diferencia:

$$\text{Temperatura ambiente: } \frac{23,5}{2,7} = 8,70$$

$$\text{Temperatura alta: } \frac{75,4}{20,9} = 3,61$$

A temperatura ambiente la CTX produce casi un incremento de nueve veces en las aberraciones cromosómicas, mientras que a alta temperatura el incremento es inferior a cuatro veces. Por tanto, en forma relativa, el efecto de la CTX es mucho mayor a temperatura ambiente. ■

Si se piensa que el fenómeno bajo estudio es multiplicativo en vez de aditivo, de forma que el interés principal es el cambio relativo en vez de absoluto, entonces no se deben utilizar contrastes ordinarios. Un planteamiento simple en esta situación es utilizar una transformación logarítmica (es decir, calcular $Y' = \log(Y)$, y analizar después Y' mediante contrastes. La motivación de este planteamiento es que las relaciones de magnitud *relativa* (cocientes) constante en la escala de Y se convierten en relaciones de magnitud *absoluta* (diferencias) constante en la escala de Y' .

Ejercicios 11.8.1-11.8.10

11.8.1 En referencia a los datos de CVF del Ejemplo 11.8.1:

- (a) Verifique que la media global de los 481 valores de CVC es 4,56.
- (b) Teniendo en cuenta la distribución de edades de los 481 sujetos y la distribución de edades en la población de Estados Unidos, explique intuitivamente por qué la media global (4,56 litros) es menor que la media ajustada por edad (4,67 litros).

11.8.2 Para ver si hay una relación entre la presión sanguínea y la maternidad, unos investigadores examinaron datos de un gran estudio sobre la salud. La tabla siguiente muestra los datos de presión sanguínea sistólica (mmHg) para muestras aleatorias de dos poblaciones de mujeres: mujeres que no habían tenido hijos y mujeres que habían tenido cinco o más hijos. La desviación típica conjunta de los ocho grupos es $s_{\text{conjunta}} = 18 \text{ mmHg}^{25}$.

Edad	Sin hijos		Cinco o más hijos	
	Presión sanguínea media	Número de mujeres	Presión sanguínea media	Número de mujeres
18-24	113	230	114	7
25-34	118	110	116	82
35-44	125	105	124	127
45-54	134	123	138	124
18-54	121	568	127	340

Realice el ajuste por edad, en la forma explicada, utilizando la siguiente distribución de referencia, que es la distribución aproximada de las mujeres estadounidenses²⁶.

Edad	Frecuencia relativa
18-24	0,17
25-34	0,29
35-44	0,31
45-54	0,23

- (a) Calcule la presión sanguínea media ajustada por edad de las mujeres sin hijos.
- (b) Calcule la presión sanguínea media ajustada por edad de mujeres con cinco o más hijos.

- (c) Calcule la diferencia entre los valores obtenidos en los apartados (a) y (b). Explique intuitivamente por qué el resultado es menor que la diferencia no ajustada de $127 - 121 = 6 \text{ mmHg}$.
- (d) Calcule el error típico del valor calculado en el apartado (a).
- (e) Calcule el error típico del valor calculado en el apartado (c).

11.8.3 En referencia a los datos de ATP del Ejercicio 11.71, las medias muestrales y las desviaciones típicas son las siguientes:

	Abedul de río		Abedul europeo	
	Inundación	Control	Inundación	Control
$\bar{y}$	1,19	1,78	0,29	1,20
s	0,18	0,24	0,20	0,16

Defina combinaciones lineales (es decir, especifique los coeficientes) para medir lo siguiente:

- (a) El efecto de la inundación en el abedul de río.
- (b) El efecto de la inundación en el abedul europeo.
- (c) La diferencia entre el abedul de río y el abedul europeo con respecto al efecto de la inundación (es decir, la interacción entre la inundación y las especies).

11.8.4 (Continuación del Ejercicio 11.8.3)

- (a) Utilice un test t para investigar si la inundación tiene el mismo efecto en el abedul del río que en el abedul europeo. Utilice una alternativa no direccional y $\alpha = 0,05$. (La desviación típica conjunta es $s_{\text{conjunta}} = 0,199$).
- (b) Si los tamaños muestrales fueran $n = 10$ en vez de $n = 4$ en cada grupo, pero las medias, las desviaciones típicas y los valores de s_{conjunta} permanecieron iguales, ¿cómo cambiaría el resultado del apartado (a)?

11.8.5 (Continuación del Ejercicio 11.8.4)

Considere la hipótesis nula de que la inundación no tiene efecto en los niveles de ATP del abedul de río. Esta hipótesis se podría probar de dos formas: mediante un contraste (utilizando el método de la Sección 11.8) o mediante un test t de dos muestras (como en el Ejercicio 7.2.11). Responda a las siguientes preguntas; no es necesario que realice los contrastes.

- (a) ¿En qué forma o formas se diferencian los dos procedimientos de contraste?
- (b) ¿En qué forma o formas se diferencian las condiciones de validez de los dos procedimientos?

- (c) Uno de los dos procedimientos requiere más condiciones para su validez; pero si esas condiciones se cumplen, entonces dicho procedimiento presenta ventajas sobre el otro. ¿Cuáles son esas ventajas?

11.8.6 Considere los datos del Ejercicio 11.7.3, donde se compararon las sustancias ticrinafeno (T) e hidroclorotiazida (H). Los datos se resumen en la tabla siguiente. La desviación típica conjunta es $s_{\text{conjunta}} = 11,83$ mmHg.

	Ticrinafeno (T)		Hidroclorotiazida (H)	
	Dosis baja	Dosis alta	Dosis baja	Dosis alta
Media	13,9	17,1	15,8	17,5
Número de pacientes	53	57	55	58

Si las dos sustancias tienen efectos iguales sobre la presión sanguínea, entonces es preferible el ticrinafeno porque tiene menos efectos secundarios.

- (a) Construya un intervalo de confianza del 95 % para la diferencia entre las sustancias (con respecto a la reducción media de la presión sanguínea), promediada sobre los dos niveles de dosificación.
- (b) Interprete el intervalo de confianza del apartado (a) en el contexto de este ejercicio.

11.8.7 Considere el experimento sobre crecimiento de lechugas descrito en el Ejercicio 11.7.6. La tabla que acompaña a este ejercicio muestra el peso medio en seco de las hojas (g) de nueve plantas en cada grupo de tratamiento. El valor de CM(intra) del ANOVA fue 0,3481.

	Solución de nutrientes	
	Estándar	Nitrógeno extra
Iluminación baja	2,16	3,09
Iluminación alta	3,26	4,48

Construya un intervalo de confianza del 95 % para el efecto del nitrógeno, promediado sobre las dos condiciones de iluminación.

11.8.8 En referencia a los datos de MAO del Ejercicio 11.4.1:

- (a) Defina un contraste para comparar la actividad de la MAO para esquizofrénicos sin características paranoi-

cas frente al promedio de los dos tipos de características paranoicas.

- (b) Calcule el valor del contraste del apartado (a) y su error típico.
- (c) Aplique un test t al contraste del apartado (a). Sea H_A la alternativa no direccional y $\alpha = 0,05$.

11.8.9 ¿Son anatómicamente diferentes los cerebros de los zurdos? Para investigar esta cuestión, un neurólogo realizó un examen post mórtem del cerebro de 42 personas. Cada una de ellas había sido evaluada antes de fallecer con respecto a sus preferencias de uso de la mano, y clasificadas como consistentemente diestros (CD) o mixtos (M). La tabla que sigue muestra los resultados del área de la mitad anterior del cuerpo calloso (la estructura que enlaza los hemisferios derecho e izquierdo del cerebro)²⁷. El valor de CM(intra) del ANOVA fue 2,498.

Grupo	Área (mm ²)		
	Media	DT	n
1. Hombres: M	423	48	5
2. Hombres: CD	367	49	7
3. Mujeres: M	377	63	10
4. Mujeres: CD	345	43	20

- (a) La diferencia entre M y CD es de 56 mm² para los hombres y de 32 mm² para las mujeres. ¿Es esta evidencia suficiente para concluir que la correspondiente diferencia poblacional es mayor para los hombres que para las mujeres? Contraste la hipótesis apropiada. (Use una alternativa no direccional y $\alpha = 0,10$).
- (b) Como medida global de la diferencia entre M y CD, se puede considerar el valor $0,5(\mu_1 - \mu_2) + 0,5(\mu_3 - \mu_4)$. Construya un intervalo de confianza del 95 % para este valor. (Se trata de una comparación ajustada por sexo de M y CD, donde la población de referencia es el 50 % de hombres y el 50 % de mujeres).

11.8.10 Considere los datos sobre narcisos del Ejercicio 11.4.5.

- (a) Defina un contraste para comparar la longitud del tallo de los narcisos del área abierta frente al promedio de los lados norte, sur, este y oeste del edificio.
- (b) Calcule el valor del contraste del apartado (a) y su error típico.
- (c) Aplique un test t al contraste del apartado (b). Considere H_A como alternativa no direccional y $\alpha = 0,05$.

11.9 Comparaciones múltiples (opcional)

Tras realizar nuestro test F global, podemos obtener que existe evidencia significativa a favor de una diferencia entre las medias poblacionales $\mu_1, \mu_2, \dots, \mu_I$. En esta situación, a menudo estaremos interesados en realizar un análisis detallado de las medias muestrales $\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_I$ considerando todas las comparaciones por parejas. Es decir, deseamos contrastar todas las posibles hipótesis por parejas:

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_1 = \mu_3$$

$$H_0: \mu_2 = \mu_3$$

y así sucesivamente.

Vimos en la Sección 11.1 que la utilización de tests t repetidos conduce a un incremento del riesgo global de cometer un error de Tipo I (es decir, obtener evidencia de una diferencia en las medias poblacionales cuando, de hecho, no hay diferencia). De hecho, fue este aumento en el riesgo de cometer un error de Tipo I lo que motivó la realización en primer lugar del test F . En esta sección describiremos tres métodos de comparación múltiple para controlar el riesgo global de cometer un error de Tipo I: el método de Bonferroni, la Diferencia Significativa Mínima de Fisher y la Diferencia Significativa Honesta de Tukey. No obstante, debemos examinar primero los diferentes tipos de error de Tipo I que pueden aparecer cuando se consideran comparaciones múltiples.

Error respecto a experimentos frente a error respecto a comparaciones

Consideremos un estudio que requiere la comparación de cuatro medias poblacionales: μ_1, μ_2, μ_3 y μ_4 . Como se indicó en la Sección 11.1, hay cuatro comparaciones posibles:

$$H_0: \mu_1 = \mu_2 \quad H_0: \mu_1 = \mu_3 \quad H_0: \mu_1 = \mu_4 \quad H_0: \mu_2 = \mu_3 \quad H_0: \mu_2 = \mu_4 \quad H_0: \mu_3 = \mu_4$$

Al considerar las seis comparaciones, podemos hablar de la probabilidad de cometer un error de Tipo I en una comparación particular, por ejemplo $H_0: \mu_1 = \mu_2$, que se denomina **probabilidad de error de Tipo I respecto a comparaciones** (α_{cw}), o podemos hablar de la probabilidad de cometer un error de Tipo I en *cualquiera* de las seis comparaciones, que se denomina **probabilidad de error de Tipo I con respecto al experimento** (α_{ew})*. Por ejemplo, la Tabla 11.1.2 muestra las probabilidades de error de Tipo I respecto al experimento al comparar un número diferente de grupos cuando la probabilidad de error de Tipo I respecto al experimento es $\alpha_{cw} = 0,05$.

Aunque la relación entre α_{cw} y α_{ew} puede ser compleja, siempre se cumple que

$$\alpha_{ew} \leq k \times \alpha_{cw}$$

siendo k el número de comparaciones. Por tanto, si se realizan seis comparaciones independientes con un nivel $\alpha_{cw} = 0,05$, la probabilidad de error respecto al experimento (α_{ew}) es como mucho $6 \times 0,05 = 0,30$.

Diferencia significativa mínima de Fisher

En la Sección opcional 11.8 presentamos un procedimiento para estimar contrastes lineales. La Diferencia Significativa Mínima de Fisher (DSM) usa este procedimiento para producir intervalos de confianza por parejas para diferencias poblacionales utilizando $\alpha_{cw} = \alpha$, el error de Tipo I utilizado en el ANOVA. Los intervalos que no contienen al cero proporcionan evidencia de una diferencia significativa entre las medias poblacionales comparadas.

Se presenta a continuación un ejemplo del procedimiento.

Ejemplo 11.9.1 Ostras y pasto marino

En un estudio para investigar el efecto de la densidad de ostras en la biomasa de los pastos marinos, unos investigadores pusieron ostras en 30 parcelas de 1 m² de pasto marino sano. Al principio del estudio el pasto marino se cortó dejándolo con poca longitud en todas las parcelas. Seguidamente, 10 parcelas elegidas aleatoriamente recibie-

* Aunque el término *respecto al experimento* contiene la palabra experimento, esta terminología se emplea tanto en estudios experimentales como observacionales.

ron una alta densidad de ostras; otras 10 una densidad intermedia y otras 10 una densidad baja. Como control, 10 parcelas de pasto marino cortado y de 1 m² no recibieron ostras. Tras dos semanas, se midió la biomasa del pasto marino bajo tierra en cada parcela (g/m²). No se pudieron medir los datos de algunas parcelas. Se presentan a continuación el resumen de los datos (Tabla 11.9.1) y la tabla del ANOVA (Tabla 11.9.2)²⁸.

Tabla 11.9.1 Biomasa de pasto marino bajo tierra (g/m ²)				
	Densidad de ostras			
	Ninguna (1)	Baja (2)	Intermedia (3)	Alta (4)
Media	34,81	33,13	28,33	15,00
DT	13,44	17,36	17,11	10,97
<i>n</i>	9	10	8	10

Tabla 11.9.2 Resumen del ANOVA para la biomasa de pasto marino bajo tierra (g/m ²)					
	gl	Suma de cuadrados	Cuadrados medios	<i>F</i>	<i>P</i> valor
Entre	3	2.365,5	788,51	3,5688	0,0243
Dentro de	33	7.291,1	220,94		
Total	36	9.656,6			

El *P* valor del ANOVA es 0,0243, lo que indica que hay evidencia significativa a favor de una diferencia entre las biomásas medias bajo estas condiciones experimentales. Una vez tenemos evidencia de la diferencia, procedemos con las comparaciones.

Recordemos que para cualquier contraste lineal $L = m_1\bar{y}_1 + m_2\bar{y}_2 + \dots + m_I\bar{y}_I$,

$$ET_L = s_{\text{conjunta}} \sqrt{\sum_{i=1}^I \frac{m_i^2}{n_i}}$$

siendo

$$s_{\text{conjunta}} = \sqrt{\text{CM}(\text{intra})}$$

Por tanto, para comparar la condición de sin ostras (1) con la condición de baja densidad de ostras (2), definimos $D_{12} = \bar{Y}_1 - \bar{Y}_2$, de manera que planteando un contraste lineal tenemos

$$\begin{aligned} d_{12} &= 1\bar{y}_1 + (-1)\bar{y}_2 + 0\bar{y}_3 + 0\bar{y}_4 \\ &= (1)(34,81) + (-1)(33,13) + (0)(28,33) + (0)(15,00) \\ &= 34,81 - 33,13 = 1,68 \end{aligned}$$

y, como $s_{\text{conjunta}} = \sqrt{220,94} = 14,86$, tenemos

$$\begin{aligned} ET_{D_{12}} &= 14,86 \times \sqrt{\frac{1^2}{9} + \frac{(-1)^2}{10} + \frac{0^2}{8} + \frac{0^2}{10}} \\ &= 14,86 \times \sqrt{\frac{1}{9} + \frac{1}{10}} \\ &= 6,82 \end{aligned}$$

El intervalo de confianza del 95 % para la diferencia de medias poblacionales de la biomasa bajo tierra bajo la condición de no ostras comparada con la condición de baja densidad de ostras, $\mu_1 - \mu_2$, es

$$d_{12} \pm t_{33, 0,025} \times ET_{D_{12}} = 1,68 \pm 2,0345 \times 6,82$$

$$= 1,68 \pm 13,89, \text{ es decir } (-12,21, 15,57)$$

Tenemos una confianza del 95 % en que la biomasa media bajo tierra cuando no hay ostras sea entre 12,21 g/m² inferior y 15,57 g/m² superior a la biomasa cuando hay una densidad baja de ostras. Como el intervalo contiene al cero, no hay evidencia de que la biomasa media bajo tierra sea diferente bajo esas dos condiciones.

Repetiendo el proceso para las restantes cinco comparaciones se obtienen los cálculos intermedios y los resultados finales que se muestran en la Tabla 11.9.3.

Tabla 11.9.3 Cálculos intermedios e intervalos basados en la DSM de Fisher comparando la biomasa bajo tierra en función de diferentes valores de la densidad de ostras*				
Comparación	$d_{ab} = \bar{y}_a - \bar{y}_b$	$\sqrt{(1/n_a) + (1/n_b)}$	$ET_{D_{ab}} = s_{conjunta} \times \sqrt{(1/n_a) + (1/n_b)}$	$t_{33, 0,025} \times ET_{D_{ab}}$
Ninguna-baja	1,68	0,459	6,828	13,891
Ninguna-intermedia	6,48	0,486	7,221	14,690
<i>Ninguna-alta</i>	<i>19,81</i>	<i>0,459</i>	<i>6,828</i>	<i>13,891</i>
Baja-intermedia	4,80	0,474	7,049	14,341
<i>Baja-alta</i>	<i>18,13</i>	<i>0,447</i>	<i>6,646</i>	<i>13,520</i>
Intermedia-alta	13,33	0,474	7,049	14,341
Comparación	Inferior 95 %	Superior 95 %		
Ninguna-baja	− 12,2	15,6		
Ninguna-intermedia	− 8,2	21,2		
<i>Ninguna-alta</i>	<i>5,9</i>	<i>33,7</i>		
Baja-intermedia	− 9,5	19,1		
<i>Baja-alta</i>	<i>4,6</i>	<i>31,7</i>		
Intermedia-alta	− 1,0	27,7		
* Los intervalos que no contienen a cero (es decir, una diferencia estadísticamente significativa entre las media de grupo) están en cursiva. Nótese que un intervalo no contendrá a cero siempre que $ D_{ab} > t \times ET_{D_{ab}}$. (El valor de $t_{33, 0,025} = 2,0345$ se ha determinado mediante computador. Utilizando la Tabla 4 se obtendrían resultados muy similares con el valor de la tabla correspondiente a 30 grados de libertad, $t_{30, 0,025} = 2,042$).				

Observando la Tabla 11.9.3 podemos ver que las únicas comparaciones que muestran diferencias significativas en la biomasa media son la de sin ostras con alta densidad de ostras y la de baja con alta densidad de ostras. ■

Una fórmula general para calcular un intervalo de DSM de Fisher con confianza $100(1 - \alpha) \%$ para $(\mu_a - \mu_b)$ se da en el recuadro siguiente.

Intervalo de confianza del $100(1 - \alpha) \%$ de DSM de Fisher para $(\mu_a - \mu_b)$

siendo

$$d_{ab} \pm t_{gl, \alpha/2} \times ET_{D_{ab}}$$

$$d_{ab} = \bar{y}_a - \bar{y}_b$$

$$ET_{D_{ab}} = s_{conjunta} \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}$$

$$s_{conjunta} = \sqrt{CM(intra)}$$

$$gl = gl(intra)$$

y

¿Cómo controla la DSM de Fisher el error de Tipo I con respecto al experimento? Las comparaciones del DSM de Fisher se deben realizar después de rechazar la hipótesis nula global del ANOVA, correspondiente a que todas las medias poblacionales son iguales: $H_0: \mu_1 = \mu_2 = \dots = \mu_I$. El test F global del ANOVA actúa como un procedimiento de comprobación de las comparaciones múltiples y, por tanto, ofrece control sobre α_{ew} .

Presentación de resultados

La presentación de los seis intervalos de DSM de Fisher para el ejemplo del pasto marino de la Tabla 11.9.3 es un resumen de trabajo útil, pero no es adecuada para una comunicación efectiva de resultados. Para organizar los resultados en una presentación con una tabla simple seguiremos los siguientes pasos.

Paso 1 *Agrupación de las etiquetas de grupo.* Dispondremos las etiquetas de los grupos en orden creciente con respecto a sus medias.

Paso 2 *Comparación sistemática de medias, subrayando las comparaciones no significativas.*

- Se comienza examinando la comparación de intervalos de las medias máxima y mínima. Si el intervalo contiene al cero, la diferencia de medias no es estadísticamente significativa y se dibuja una línea debajo del conjunto de etiquetas de grupo que «conecte» los grupos con las medias máxima y mínima. Si el intervalo no contiene al cero, seguir con el paso siguiente.
- Ignorar el grupo con la media mínima y comparar los restantes subgrupos de $I - 1$ medias. Como en el paso 2(a), si el intervalo contiene al cero, la diferencia de medias no es estadísticamente significativa y se dibuja una línea debajo del conjunto de etiquetas de grupo que «conecte» los grupos. Seguidamente se considera el siguiente subconjunto de $I - 1$ medias (las medias que quedan en el grupo con la media máxima se ignoran). De nuevo, se subraya el subconjunto si la media contiene al cero.
- Se repite el paso 2(b) comparando sucesivamente subconjuntos de tamaño $I - 2$, $I - 3$, y así sucesivamente, hasta que se produzca un intervalo que contenga al cero o no se puedan realizar más comparaciones.

Notas importantes: Durante este procedimiento, nunca se realizará una comparación con cualquier subconjunto que ya haya sido subrayado. Esas medias de grupo se declaran automáticamente como sin diferencia estadísticamente significativa. Además, al subrayar, debe utilizarse una línea separada para cada paso. Nunca se debe unir una línea con otra que ya haya sido dibujada.

Paso 3 *Trasladamos los subrayados a un resumen en forma de tabla.* Crear un resumen en forma de tabla de los datos utilizando superíndices para indicar qué grupos no tienen una diferencia estadísticamente significativa.

Ejemplo 11.9.2 Ostras y pasto marino

En este ejemplo seguiremos el procedimiento anterior para presentar las comparaciones de la DSM de Fisher en el caso de las ostras y el pasto marino que se mostraron en la Tabla 11.9.3.

Paso 1 En primer lugar se disponen las etiquetas en orden creciente de medias (como muestra la Tabla 11.9.1).

Alta Intermedia Baja Ninguna

Paso 2 Comparamos los grupos con las medias mínima (alta densidad de ostras) y máxima (ninguna ostra): $\mu_{\text{Ninguna}} - \mu_{\text{Alta}} = (5,9, 33,7)$. Este intervalo no contiene al cero, por lo que esas medias son significativamente diferentes y no se subraya*. Procedemos ahora con el paso siguiente (paso 2b), las comparaciones de subconjuntos de tres medias. Primero, comparamos de Intermedia a Ninguna:

$$\mu_{\text{Ninguna}} - \mu_{\text{Intermedia}} = (-8,2, 21,2)$$

Este intervalo contiene al cero, por lo que se procede a subrayar como se indica a continuación.

Alta Intermedia Baja Ninguna

* Intuitivamente, este intervalo no debería contener al cero, ya que hemos rechazado la hipótesis nula del test F global, aunque existen algunas ocasiones en las que nuestro procedimiento de comparación múltiple y el test F global pueden no concordar.

Este subrayado indica que los tres grupos no tienen medias estadísticamente diferentes. Comparamos ahora el siguiente subconjunto de tres medias, de Alta a Baja: $\mu_{\text{Alta}} - \mu_{\text{Baja}} = (4,6, 31,7)$. Este intervalo no contiene al cero, por lo que no se subraya. Hay evidencia de diferencia en las biomazas bajo tierra medias entre las condiciones de alta y baja densidad de ostras.

Habiendo comparado todos los subconjuntos de tres medias, continuamos ahora con los subconjuntos de dos medias. El único subconjunto de dos medias no conectado por un subrayado es la comparación Alta-Intermedia. Este intervalo $\mu_{\text{Intermedia}} - \mu_{\text{Alta}} = (-1,0, 27,7)$ contiene al cero, por lo que se subraya como se indica a continuación.

Alta Intermedia Baja Ninguna

Paso 3 Comunicando estos resultados, otorgamos una letra a cada línea y mostramos estas letras como superíndices en nuestra tabla de medias de grupo como se muestra a continuación y en la Tabla 11.9.4. También es posible elaborar una muestra gráfica como la que desarrolla la Figura 11.9.1.

Alta Intermedia Baja Ninguna
a _____
b _____

Tabla 11.9.4 Biomasa de pasto marino bajo tierra (g/m ²) para diversos niveles de densidad de ostras*				
	Densidad de ostras			
	Ninguna	Baja	Intermedia	Alta
Media	34,8 ^a	33,1 ^a	28,3 ^{a, b}	15,0 ^b
DT	13,4	17,4	17,1	11,0
n	9	10	8	10
* Los grupos que comparten una superíndice común tienen medias que no tienen diferencia estadísticamente significativa, basándose en las comparaciones de la DSM de Fisher con $\alpha_{cw} = 0,05$.				

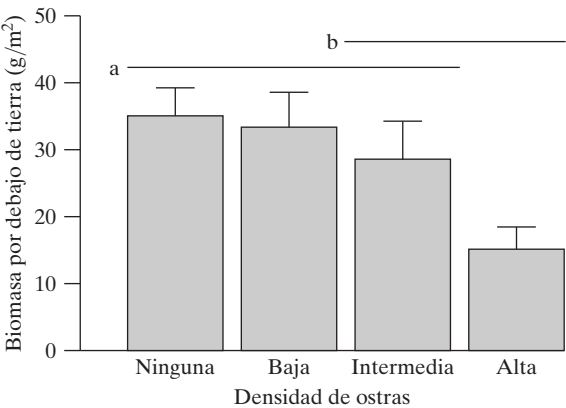


Figura 11.9.1 Biomasa de pasto marino por debajo del suelo (g/m²) para diferentes niveles de densidad de ostras. Las barras representan las medias más el error típico. Los grupos que comparten una línea superior común no presentan diferencias estadísticamente significativas de acuerdo con las comparaciones de la DSM de Fisher con $\alpha_{cw} = 0,05$

El método de Bonferroni

El **método de Bonferroni** se basa en una relación muy simple y general: la probabilidad de que ocurra al menos uno de varios sucesos no puede superar a la suma de las probabilidades individuales. Por ejemplo, supongamos que realizamos seis contrastes de hipótesis, cada uno de ellos con $\alpha_{cw} = 0,01$. Entonces, el riesgo total de error de Tipo I α_{ew} (es decir, la probabilidad de rechazar al menos una de las seis hipótesis cuando de hecho todas ellas son ciertas) no puede superar el valor

$$0,01 + 0,01 + 0,01 + 0,01 + 0,01 + 0,01 = (6)(0,01) = 0,06$$

Invirtiéndola esta lógica, supongamos que un investigador planea realizar seis contrastes de hipótesis y desea que el riesgo total de cometer un error de Tipo I no supere el valor de $\alpha_{ew} = 0,05$. Un planteamiento conservativo es realizar cada uno de los tests individuales con el nivel de significación de $\alpha_{cw} = 0,05/6 = 0,0083$. Esto se denomina **ajuste de Bonferroni**.

Nótese que la técnica de Bonferroni es de aplicación muy amplia. Los tests individuales pueden relacionarse con diferentes variable de respuesta, diferentes subconjuntos, etc. Algunos pueden ser tests t , otros tests chi-cuadrado, etc.

El planteamiento de Bonferroni puede ser empleado por una persona que está leyendo un informe de investigación, si su autor incluye P valores explícitos. Por ejemplo, si el informe contiene seis P valores y el lector desea un nivel de protección global del 5 % contra el error de Tipo I, entonces el lector no considerará un P valor como de evidencia suficiente de un efecto a menos que sea menor que $\alpha_{cw} = 0,0083$.

El ajuste de Bonferroni se puede realizar también para intervalos de confianza. Por ejemplo, supongamos que deseamos construir seis intervalos de confianza y deseamos una probabilidad global del 95 % de que *todos* los intervalos contengan sus respectivos parámetros ($\alpha_{ew} = 0,05$). Entonces esto se puede conseguir construyendo cada intervalo con un nivel de confianza del 99,17 % (ya que $0,05/6 = 0,0083$ y $1 - 0,0083 = 0,9917$).

En general, para construir k intervalos de confianza ajustados por Bonferroni con una probabilidad global de $100(1 - \alpha_{ew})$ % de que *todos* los intervalos contengan a sus respectivos parámetros, construiremos cada intervalo con un nivel de confianza de $100(1 - \alpha_{cw})$ %, siendo $\alpha_{cw} = \alpha_{ew}/k$. La mecánica de los cálculos es idéntica a la utilizada en el método de la DSM de Fisher, excepto porque el valor del multiplicador t se modifica: $t_{g1, \alpha_{cw}/2}$. Nótese que la aplicación de esta idea requiere el uso de valores críticos inusuales, por lo que no son suficientes las tablas estándar. La Tabla 11 (al final del libro) proporciona los multiplicadores de Bonferroni para intervalos de confianza basados en una distribución t . También se puede emplear *software* estadístico para obtener multiplicadores apropiados. El Ejemplo 11.9.3 ilustra esta idea.

Ejemplo 11.9.3 Ostras y pasto marino

Para calcular los intervalos de confianza del 95 % relacionados con el experimento y ajustados por Bonferroni en nuestro ejemplo de ostras y pasto marino, recordemos primero que se requieren un total de seis comparaciones, por lo que $\alpha_{cw} = 0,05/6 = 0,0083$ y $t_{30, 0,0083/2} = 2,825$ [como la Tabla 12 no contiene todos los valores de grados de libertad, usamos $gl = 30$, el valor más cercano a $gl(\text{intra}) = 33$]. La Tabla 11.9.5 resume el conjunto de intervalos de una forma similar a los intervalos del método de la DSM de Fisher de la Tabla 11.9.3.

Tabla 11.9.5 Cálculos intermedios e intervalos de Bonferroni del 95 % respecto al experimento (99,17 % respecto a las comparaciones) comparando la biomasa bajo tierra en función de diferentes valores de la densidad de ostras*

Comparación	$d_{ab} = \bar{y}_a - \bar{y}_b$	$ET_{D_{ab}}$	$t_{30, 0,025/6} \times ET_{D_{ab}}$	Inferir 99,17 %	Superior 99,17 %
Ninguna-baja	1,68	6,828	13,891	-17,6	21,0
Ninguna-intermedia	6,48	7,221	14,690	-13,9	26,9
<i>Ninguna-alta</i>	<i>19,81</i>	<i>6,828</i>	<i>13,891</i>	<i>0,5</i>	<i>39,1</i>
Baja intermedia	4,80	7,049	14,341	-15,1	24,7
Baja-alta	18,13	6,646	13,520	-0,6	36,9
Intermedia-alta	13,33	7,049	14,341	-6,6	33,2

* Los intervalos que no contienen a cero (es decir, hay una diferencia estadísticamente significativa entre las medias del grupo) están en cursiva. Nótese que las dos primeras columnas (d_{ab} y $ET_{D_{ab}}$) son idénticas a las presentadas en la Tabla 11.9.3.

Utilizando el método del subrayado para visualizar las comparaciones, tenemos

Alta Intermedia Baja Ninguna
 a _____
 b _____

Los subrayados indican que la única diferencia significativa en la biomasa media de pasto marino bajo el suelo es entre las condiciones de alta densidad de ostras y ninguna ostra. La Tabla 11.9.6 presenta un resumen de los resultados. ■

Tabla 11.9.6 Biomasa de pasto marino bajo tierra (g/m ²) para diversos niveles de densidad de ostras*				
	Densidad de ostras			
	Ninguna	Baja	Intermedia	Alta
Media	34,8 ^a	33,1 ^{a, b}	28,3 ^{a, b}	15,0 ^b
DT	13,4	17,4	17,1	11,0
<i>n</i>	9	10	8	10
* Los grupos que comparten una superíndice común tienen medias que no tienen diferencia estadísticamente significativa, basándose en las comparaciones de Bonferroni con $\alpha_{cw} = 0,05$.				

Nótese que los intervalos del método de la DSM de Fisher y de Bonferroni no son idénticos (los de Bonferroni son más amplios debido al menor valor de α_{cw}). Adicionalmente, las conclusiones también son diferentes. Los intervalos del método de la DSM de Fisher indican que hay evidencia a favor de que las condiciones de alta y baja densidad de ostras tienen diferentes medias poblacionales, mientras que el método de Bonferroni no indica esa diferencia. Esto es debido a que los intervalos de Bonferroni son menos potentes y, por tanto, más conservadores que los intervalos de Fisher. A diferencia de los intervalos de Fisher, los intervalos de Bonferroni garantizan que se obtiene una α_{cw} menor o igual que la probabilidad de error de Tipo I relacionada con el experimento deseado.

Desafortunadamente, los intervalos de Bonferroni son a menudo excesivamente conservadores, de forma que el valor real de α_{cw} es mucho menor que la probabilidad de error de Tipo I relacionada con el experimento deseado y, por tanto, se sacrifica mucha potencia en aras de la protección contra el error de Tipo I. Un procedimiento más complejo que (cuando los tamaños muestrales son iguales) es capaz de conseguir exactamente la probabilidad de error asociada al experimento (y, por tanto, conseguir una potencia mayor que Bonferroni) es el método de la Diferencia Significativa Honesta de Tukey.

La Diferencia Significativa Honesta de Tukey

El método de la Diferencia Significativa Honesta de Tukey es muy similar a los métodos de la DSM de Fisher y de Bonferroni, pero en vez de utilizar multiplicadores *t* en las fórmulas del intervalo de confianza, se utilizan valores relacionados provenientes de una distribución denominada distribución de rangos «studentizada». La mayoría de los paquetes estadísticos calculan todos los intervalos por parejas del método de la DSH de Tukey para cualquier probabilidad de error α_{cw} relacionada con el experimento. Como ejemplo, la Figura 11.9.2 muestra la salida del método de Tukey del paquete de *software* estadístico R utilizando los datos de ostras y pasto marino. Nótese que además de los intervalos, la

	diff	lwr	upr	p-adj
int-high	13.33	-5.74	32.40	0.2515
low-high	18.13	0.15	36.11	0.0475
no-high	19.81	1.34	38.28	0.0318
low-int	4.80	-14.27	23.87	0.9037
zero-int	6.48	-13.06	26.02	0.8063
zero-low	1.68	-16.79	20.15	0.9947

Figura 11.9.2 Salida del *software* R presentando los intervalos del 95 % de confianza del método de DSH de Tukey para el ejemplo de las ostras y el pasto marino

mayoría del *software* proporciona también un P valor «ajustado». Incluso realizando comparaciones múltiples, si esos P valores «ajustados» se comparan con α_{ew} , se mantendrá todavía una probabilidad de error de Tipo I asociada al experimento de α_{ew} .

Los intervalos de la Figura 11.9.2 muestran que las conclusiones extraídas del método de la DSH de Tukey coinciden con las de los intervalos obtenidos mediante el método de la DSM de Fisher: las densidades de ostras alta y baja, así como la alta y ninguna, difieren significativamente. Los extremos de los intervalos de confianza relacionados con el experimento del método de la DSH de Tukey son, sin embargo, diferentes tanto de los del método de la DSM de Fisher como de los del método de Bonferroni.

Condiciones de validez

Los tres procedimientos de comparación múltiple descritos requieren que se cumplan las mismas condiciones estándar del ANOVA dadas en la Sección 11.5. Además, las condiciones de validez para el método de la DSM de Fisher requieren también que no se utilice el procedimiento a menos que se haya rechazado la hipótesis de que todas las medias son iguales. Por el contrario, el método de la DSH de Tukey y el de Bonferroni no requieren que se realice el test F global a priori (aunque es necesario el cálculo de la $s_{conjunta}$). Para conseguir exactamente la probabilidad de error de Tipo I relacionada con el experimento, el método de la DSH de Tukey requiere que todas las muestras sean del mismo tamaño. Si los tamaños muestrales no son iguales, la probabilidad de error real será algo menor que la nominal, produciéndose una pérdida de potencia.

Una ventaja del método de Bonferroni es que es ampliamente aplicable y se puede generalizar a situaciones más allá del ANOVA. Un ejemplo de esto se presenta en los ejercicios.

Ejercicios 11.9.1-11.9.8

11.9.1 Un botánico utilizó un diseño completamente aleatorizado para asignar 45 tiestos con plantas de berenjena a cinco tratamientos diferentes del suelo. La variable observada fue el peso en seco total de la planta sin raíces (g) tras 31 días de crecimiento. Las medias de los tratamientos fueron las que se muestran en la tabla siguiente²⁹. El CM(intra) fue 0,2246. Utilice los intervalos de DSM de Fisher para comparar todas las parejas de medias con $\alpha_{cw} = 0,05$. Presente sus resultados en una tabla resumen similar a la Tabla 11.9.4. (*Sugerencia:* tenga en cuenta que todos los tamaños muestrales son iguales; por tanto, solo se necesita obtener una vez el margen de error calculado en todas las comparaciones. Hay un total de 10 comparaciones posibles).

Tratamiento	A	B	C	D	E
Media	4,37	4,76	3,70	5,41	5,38
n	9	9	9	9	9

11.9.2 Repita el Ejercicio 11.9.1, pero utilice los intervalos de Bonferroni con $\alpha_{ew} = 0,05$.

11.9.3 En un estudio del tratamiento de la anemia del ganado mediante la dieta, unos investigadores dividieron aleatoriamente a 144 vacas en cuatro grupos de tratamiento. El grupo A fue el grupo de control y los grupos B, C y D recibieron diferentes regímenes de suplementos de selenio con su dieta. Tras un año de tratamiento, se extrajeron muestras de sangre y se estudió el selenio. La tabla que acompaña a

este ejercicio muestra las concentraciones medias de selenio ($\mu\text{g/dl}$)³⁰. El CM(intra) del ANOVA fue 2,071.

Grupo	Media	n
A	0,8	36
B	5,4	36
C	6,2	36
D	5,0	36

- Calcule los intervalos ajustados por Bonferroni para comparar las dietas B, C y D con el control (dieta A) utilizando $\alpha_{ew} = 0,05$. (*Nota:* esto es un ejemplo de una situación en la que las comparaciones de Bonferroni serían preferibles a las comparaciones del DSH de Tukey, ya que no se consideran todas las comparaciones, sino que solo estamos interesados en comparar el control con cada uno de los tres tratamientos).
- En el contexto de este problema, interprete el intervalo de Bonferroni calculado en el apartado (a) que compara el control (grupo A) con el grupo que sea más diferente de dicho control.

11.9.4 Considere el experimento y los datos del Ejercicio 11.9.3. Se muestran los intervalos del método de la DSH de Tukey calculados utilizando el paquete de *software* estadístico R.

	diff	lwr	upr
B-A	4.6	3.72	5.48
C-A	5.4	4.52	6.28
D-A	4.2	3.32	5.08
C-B	0.8	-0.08	1.68
D-B	-0.4	-1.28	0.48
D-C	-1.2	-2.08	-0.32

- (a) Utilizando la salida anterior para fundamentar su respuesta, ¿existe evidencia de que cada uno de los grupos/ dietas B, C y D se diferencia del control, A?
- (b) De acuerdo con los intervalos anteriores del método DSH de Tukey y el resumen de los datos del Ejercicio 11.9.3, la dieta C produce la máxima concentración media de selenio, y es significativamente mayor que la del control. Si el objetivo de los investigadores es obtener una dieta que maximice la concentración de selenio, ¿es la dieta C la opción clara? Es decir, ¿se deberían descartar la dieta B, la D o ambas? Haga referencia a los intervalos del método DSH de Tukey en su respuesta.

11.9.5 Se comparó el efecto de diez tratamientos sobre el hígado de ratones. Había 13 animales en cada grupo de tratamiento. Los valores del ANOVA fueron $CM(intra) = 0,5842$. Los pesos medios del hígado se presentan en la tabla³¹.

Tratamiento	Peso medio del hígado (g)	Tratamiento	Peso medio del hígado (g)
1	2,59	6	2,84
2	2,28	7	2,29
3	2,34	8	2,45
4	2,07	9	2,76
5	2,40	10	2,37

- (a) Utilice los intervalos de DSM de Fisher para comparar todas las parejas de medias con $\alpha_{cw} = 0,05$ y resuma los resultados en una tabla similar a la Tabla 11.9.4. [Sugerencias para ahorrar tiempo: note primero que los tamaños muestrales son iguales, por lo que se puede utilizar el mismo margen de error ($t \times ET_{D_{ab}}$) en todas las comparaciones. Además, como se desea realizar una tabla resumen, no es necesario calcular los intervalos reales, sino solo comprobar si $|d_{ab}| > t \times ET_{D_{ab}}$. Si eso se cumple, entonces el intervalo calculado no contiene al cero, por lo que la diferencia es significativa. Finalmente, note que no es necesario realizar todas las posibles comparaciones (hay 45). Cuando se utiliza el método del subrayado para resumir resultados, una vez se ha subrayado un subconjunto de los grupos, todas las com-

paraciones dentro del subconjunto se consideran no significativas].

- (b) Si se utilizara el método de Bonferroni con $\alpha_{ew} = 0,05$ en vez del método del DSM de Fisher del apartado (a), ¿habría alguna pareja de medias significativamente diferente?

11.9.6 Considere los datos del Ejemplo 11.2.1 sobre la ganancia de peso de corderos. $CM(intra)$ del ANOVA con estos datos vale 23,333. La media muestral de la dieta 2 era 15 y de la dieta 1 era 11.

- (a) Utilice el método de Bonferroni para construir un intervalo de confianza del 95 % para la diferencia de las medias poblacionales de estas dos dietas (suponiendo que se calcularán también los intervalos para las otras dos posibles comparaciones).
- (b) Suponga que la comparación del apartado (a) es la única comparación de interés (es decir, una comparación en vez de tres). ¿Cómo cambiaría el intervalo del apartado (a)? ¿Sería más ancho, más estrecho o el mismo? Explique su respuesta.

11.9.7 Como se ha mencionado en esta sección, el procedimiento de Bonferroni se puede utilizar en una variedad de circunstancias. Consideremos el ejemplo de anidamiento de chorlitos de la Sección 10.5, que compara las localizaciones de los nidos durante tres años. La distribución de porcentajes se muestra en la tabla siguiente.

Localización	Año		
	2004	2005	2006
Campo agrícola (CA)	48,8	30,2	55,3
Hábitat de perros de la pradera (PP)	39,5	60,3	25,5
Pradera de hierba (H)	11,6	9,5	19,1
Total	99,9*	100,0	99,9*

* Las sumas de los porcentajes de 2004 y 2006 difieren del 100 % debido al redondeo.

El P valor del test chi-cuadrado de estos datos era 0,007, indicando una diferencia significativa en la distribución de las localizaciones de los nidos durante esos tres años con $\alpha = 0,10$. Considerando tablas reducidas y utilizando tests chi-cuadrado para comparar las distribuciones de las localizaciones de los nidos se obtienen los siguientes P valores:

Años comparados	P valor
2004 a 2005	0,100
2004 a 2006	0,307
2005 a 2006	0,001

Utilizando un ajuste de Bonferroni para conseguir $\alpha_{ew}=0,10$, ¿para qué pareja o parejas de años existe evidencia de una diferencia significativa en las distribuciones de la localización de los nidos? Indique el valor de α_{cw} utilizado.

11.9.8 El Ejercicio 10.5.1 presentó el siguiente problema: una serie de pacientes con osteoartritis dolorosa de la rodilla se asignaron aleatoriamente en una prueba clínica a cinco tratamientos: glucosamina, condroitina, ambos, placebo o Celebrex (la terapia estándar). Un resultado que se obtuvo fue si cada paciente experimentó o no experimentó una mejora sustancial en el dolor o en la funcionalidad. Los datos se muestran en la tabla siguiente.

Tratamiento	Resultado de éxito		
	Tamaño muestral	Número	Porcentaje
Glucosamina	317	192	60,6
Condroitina	318	202	63,5
Ambos	317	208	65,5
Placebo	313	178	56,9
Celebrex	318	214	67,3

(a) Suponga que se desea comparar solo las tasas de éxito de cada uno de los tratamientos con el control (placebo) utilizando cuatro tests chi-cuadrado separados de 2×2 .

Los P valores de esas comparaciones se presentan a continuación. Utilizando un ajuste de Bonferroni con $\alpha_{ew} = 0,05$, ¿qué tratamientos se comportan de forma significativamente diferente con respecto a placebo? Indique el valor de α_{cw} utilizado.

Tratamientos comparados con el placebo	P valor
Glucosamina	0,346
Condroitina	0,088
Ambos	0,024
Celebrex	0,007

(b) El P valor del test chi-cuadrado que considera la tabla completa de 5×2 es 0,054, lo que no proporciona suficiente evidencia para demostrar cualquier diferencia entre las tasas de éxito de los cinco tratamientos tomando $\alpha = 0,05$. Explique por qué este resultado no contradice los resultados del apartado (a). [Sugerencia: ¿Cuántas comparaciones se consideran en este test chi-cuadrado en relación con el número de comparaciones del apartado (a)? Para obtener $\alpha_{ew} = 0,05$ utilizando un ajuste de Bonferroni, ¿qué valor sería necesario que tuviera α_{cw} ? ¿Qué valor tenía en el apartado (a)? ¿Cómo afecta a la potencia de cada test el realizar muchos tests con un ajuste de Bonferroni?].

11.10 Perspectiva

En el Capítulo 11 hemos presentado algunos temas estadísticos que surgen cuando se analizan datos de más de dos muestras y hemos considerado algunos métodos clásicos de análisis. En esta sección revisaremos estos temas y mencionaremos brevemente algunos métodos alternativos de análisis.

Ventajas de un planteamiento global

Recapitemos las ventajas de analizar I muestras independientes mediante un planteamiento global en vez de realizar separadamente comparaciones por parejas.

1. *Comparaciones múltiples.* En la Sección 11.1 vimos que el uso de tests t repetidos puede aumentar grandemente el riesgo global de cometer un error de Tipo I. Se puede ganar un control del error de Tipo I por el simple hecho de empezar el análisis de los datos con un test F global. Para un control más estricto del error de Tipo I están disponibles otros métodos de comparación múltiple (por ejemplo Bonferroni y DSH de Tukey), que se describen en la Sección opcional 11.9. (Nótese que el problema de las comparaciones múltiples no es exclusivo del ANOVA).
2. *Uso de la estructura en los tratamientos o grupos.* El análisis de combinaciones adecuadas de medias de grupos puede ser muy útil para interpretar los datos. Muchas de las técnicas relevantes están más allá del alcance de este libro. La presentación en las Secciones opcionales 11.7 y 11.8 da una idea de las posibilidades. En el Capítulo 12 presentaremos algunas ideas que son aplicables cuando los propios tratamientos son cuantitativos (por ejemplo, dosis).
3. *Uso de una DT conjunta.* Hemos visto que utilizar todas las variabilidades dentro de las muestras para formar una única de DT conjunta conduce a un estimador mejor de la DT poblacional y, por tanto, a un análisis más preciso. Esto es particularmente ventajoso si los tamaños muestrales (los valores de las n) son pequeños, en cuyo caso los estimadores de las DT individuales son bastante imprecisos. Por supuesto, la utilización de una DT conjunta solo

es apropiado si las DT poblacionales son iguales. Algunas veces sucede que no se puede aprovechar la asociación de las DT porque la suposición de que las DT poblacionales son iguales no es sostenible. Un planteamiento que puede ser útil en este caso es analizar una variable transformada, como $\log(Y)$. Las DT pueden ser aproximadamente similares en la escala transformada.

Otros diseños experimentales

Las técnicas de este capítulo son válidas solo para muestras independientes. Pero la idea básica (dividir la variabilidad dentro de y entre tratamientos en componentes interpretables) se puede aplicar a muchos diseños experimentales. Por ejemplo, todas las técnicas presentadas en este capítulo se pueden adaptar (mediante una modificación adecuada de cálculo del ET) para analizar datos de un experimento con más de dos factores experimentales o situaciones en las que todos o algunos factores experimentales son numéricos en vez de categóricos. Estas técnicas y otras relacionadas con ellas pertenecen a una materia amplia denominada *análisis de la varianza*, de la que solo hemos presentado una pequeña parte.

Planteamientos no paramétricos

Existen procedimientos análogos al test de Wilcoxon-Mann-Whitney para k grupos de muestras y otros tests no paramétricos (por ejemplo, el test de Kruskal-Wallis). Esos tests tienen la ventaja de no asumir distribuciones normales subyacentes. Sin embargo, muchas de las ventajas de las técnicas paramétricas (como el uso de combinaciones lineales) no se pueden trasladar fácilmente a los planteamientos no paramétricos.

Ordenación y selección

En algunas investigaciones el objetivo principal del investigador no es responder preguntas sobre las poblaciones sino simplemente *seleccionar* una o varias poblaciones «mejores». Por ejemplo, supongamos que están disponibles 10 poblaciones (stocks) de gallinas ponedoras y se desea seleccionar la población con el máximo potencial de puesta de huevos. El investigador seleccionará una muestra aleatoria de n gallinas de cada stock y observará para cada una de ellas Y = número total de huevos puestos en 500 días³². Una pregunta relevante es: ¿Qué valor debería tener n de forma que sea probable que el stock que es *realmente* el mejor (tiene μ más alto) *parezca* también el mejor (tenga $\bar{Y}$ más alto)? Esta pregunta y preguntas similares son el objeto de estudio de una rama de estadística denominada *teoría de la ordenación y selección*.

Ejercicios suplementarios 11.S.1-11.S.19

(Nota: Los ejercicios precedidos por un asterisco se refieren a secciones opcionales).

11.S.1 Considere la investigación descrita en el Ejercicio 11.4.6 donde se estudiaron 10 mujeres en una clase de ejercicio aeróbico, 10 mujeres de una clase de baile moderno y 9 mujeres de un grupo de control. Una medida que se tomó en cada mujer fue el cambio de masa libre de grasa durante el periodo de entrenamiento de 16 semanas de curso. En la tabla siguiente se muestran los estadísticos resu-

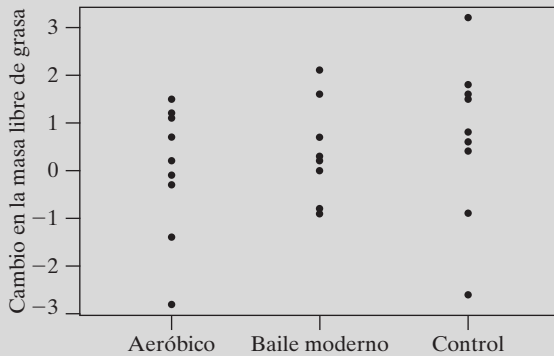
	Aeróbico	Baile moderno	Control
Media	0,00	0,44	0,71
DT	1,31	1,17	1,68
n	10	10	9

men⁸. Los valores del ANOVA de SC (entre) y SC (intra) son, respectivamente, 2,465 y 50,133.

- (a) Explique con palabras, en el contexto de este problema, la hipótesis nula contrastada por el análisis de la varianza.
- (b) Construya la tabla del ANOVA y contraste la hipótesis nula. Utilice $\alpha = 0,05$.

11.S.2 En referencia al Ejercicio 11.S.1, el test F se basa en ciertas condiciones que afectan a las distribuciones poblacionales.

- (a) Plantee las condiciones.
- (b) El diagrama de puntos de la página siguiente muestra los datos iniciales. Basándose en este diagrama y en la información dada en el Ejercicio 11.S.1, ¿parece que se cumplen las condiciones del test F ? ¿Por qué o por qué no?



11.S.3 En un estudio sobre la enfermedad ocular retinitis pigmentosa (RP), 211 pacientes se clasificaron en cuatro grupos de acuerdo con el patrón de herencia de su enfermedad. Se midió la agudeza visual (error refractivo esférico, en dioptrías) en cada ojo y después se promediaron los dos valores para obtener una única observación por persona. La tabla que acompaña a este ejercicio muestra el número de personas de cada grupo y el error refractivo medio de cada grupo³³. El ANOVA de las 211 observaciones produce $SC(\text{entre}) = 129,49$ y $SC(\text{intra}) = 2.506,8$. Construya la tabla del ANOVA y realice el test F con $\alpha = 0,05$.

Grupo	Número de personas	Error refractivo medio
RP autosómica dominante	27	+0,07
RP autosómica recesiva	20	-0,83
RP ligada al sexo	18	-3,30
RP aislada	146	-0,84
Total	211	

11.S.4 (Continuación del Ejercicio 11.S.3) Otro planteamiento para el análisis de los datos es utilizar el ojo, en vez de la persona, como unidad observacional. Las 211 personas proporcionan 422 medidas del error refractivo. La tabla que acompaña a este ejercicio resume estas medidas. En el ANOVA de las 420 observaciones resulta $SC(\text{entre}) = 258,97$ y $SC(\text{intra}) = 5.143,9$.

Grupo	Número de ojos	Error refractivo medio
RP autosómica dominante	54	+0,07
RP autosómica recesiva	40	-0,83
RP ligada al sexo	36	-3,30
RP aislada	292	-0,84
Total	422	

- Construya la tabla del ANOVA y acote el P valor del test F . Compárelo con el P valor obtenido en el Ejercicio 11.S.3. ¿Cuál de los dos P valores es de validez dudosa y por qué?
- El error refractivo medio para los pacientes de RP ligada al sexo fue de $-3,30$. Calcule el error típico de esta medida de dos formas: (i) considerando la persona como unidad observacional y utilizando del ANOVA del Ejercicio 11.S.3; (ii) considerando el ojo como unidad observacional y utilizando el ANOVA de este ejercicio. ¿Cuál de estos errores típicos es de validez dudosa y por qué?

***11.S.5** En un estudio de los efectos mutuos de los contaminantes del aire, ozono y dióxido de azufre, se cultivaron judías verdes de Blue Lake en invernaderos con el techo abierto. Algunos invernaderos se fumigaron repetidamente con dióxido de azufre. El aire de algunos invernaderos se filtró con carbón para eliminar el ozono ambiental. Había tres cámaras por combinación de tratamientos, asignadas aleatoriamente. Después de un mes de tratamiento, se midió la cosecha total de vainas (kg) en cada cámara, con los resultados que se muestran en la tabla siguiente³⁴. Para estos datos, es $SC(\text{entre}) = 1,3538$ y $SC(\text{intra}) = 0,27513$. Complete la tabla del ANOVA y realice el test F con $\alpha = 0,05$.

	Ozono ausente		Ozono presente	
	Dióxido de azufre		Dióxido de azufre	
	Ausente	Presente	Ausente	Presente
	1,52	1,49	1,15	0,65
	1,85	1,55	1,30	0,76
	1,39	1,21	1,57	0,69
Media	1,587	1,417	1,340	0,700
DT	0,237	0,181	0,213	0,056

Prepare un gráfico de interacción (como la Figura 11.7.3).

***11.S.6** Considere los datos del Ejercicio 11.S.5. Para estos datos, $SC(\text{ozono}) = 0,696$, $SC(\text{azufre}) = 0,492$, $SC(\text{interacción}) = 0,166$ y $SC(\text{intra}) = 0,275$.

- Construya la tabla del ANOVA.
- Realice un test F para las interacciones. Utilice $\alpha = 0,05$.
- Contraste la hipótesis nula de que el ozono no tiene efecto en la cosecha. Utilice $\alpha = 0,05$.

***11.S.7** En referencia al Ejercicio 11.S.5, defina contrastes para medir cada uno de los efectos especificados, y calcule el valor de cada contraste.

- El efecto del dióxido de azufre en ausencia de ozono.
- El efecto del dióxido de azufre en presencia de ozono.
- La interacción entre el dióxido de azufre y el ozono.

***11.S.8** (Continuación de los Ejercicios 11.S.6 y 11.S.7) Para los datos de las judías verdes, utilice un test t para contrastar la hipótesis nula de que no existe interacción frente a la alternativa de que el dióxido de azufre es más dañino en la presencia de ozono que en su ausencia. Utilice $\alpha = 0,05$. ¿Cómo se compara este resultado con el test F del Ejercicio 11.S.6(b) (que tiene una alternativa no direccional)?

***11.S.9 (Ejercicio para computador)** En referencia a los datos sobre las judías verdes del Ejercicio 11.S.5, aplique la transformación del inverso a los datos. Es decir, para cada valor de cosecha Y , calcule $Y' = 1/Y$.

- Calcule la tabla del ANOVA para Y' y realice el test F .
- A menudo sucede que las DT están más cerca de ser iguales en datos transformados que en datos originales. ¿Es esto cierto para los datos sobre las judías verdes cuando se utiliza la transformación del inverso?
- Realice una gráfica de probabilidad normal de los residuos, $(y'_{ij} - \bar{y}'_i)$. ¿Apoya esta gráfica la condición de que las poblaciones son normales?

***11.S.10 (Ejercicio para computador —continuación de los Ejercicios 11.S.8 y 11.S.9)** Repita el test del Ejercicio 11.S.7 utilizando Y' en vez de Y , y compare con los resultados del Ejercicio 11.S.7.

11.S.11 Suponga que se va a realizar un estudio para comparar en personas un medicamento para tratar la presión sanguínea alta con un medicamento estándar para controlar la presión sanguínea.

- Describa un diseño experimental para el estudio que utilice agrupación en bloques. Sea cuidadoso en advertir qué partes del diseño involucran aleatoriedad y qué partes no.
- ¿Puede el experimento descrito en el apartado (a) ser ciego? Si es así, explique cómo se podría hacer.

11.S.12 En un estudio de angioplastia con globo, unos pacientes con enfermedad de la arteria coronaria se asignaron aleatoriamente a cuatro grupos de tratamiento placebo, probucol (un medicamento experimental), multivitaminas (una combinación de betacaroteno, vitamina E y vitamina C), o probucol combinado con multivitaminas. Se realizó a cada uno de los pacientes con angioplastia con globo. Posteriormente, se midió en cada uno de los pacientes el «diámetro luminal mínimo» (una medida del éxito de la angioplastia para dilatar la arteria). Los estadísticos resumen se presentan en la siguiente tabla³⁵.

	Placebo	Probuco	Multi- vitaminas	Probuco y multi- vitaminas
n	62	58	54	56
Media	1,43	1,79	1,40	1,54
DT	0,58	0,45	0,55	0,61

- Complete la tabla del ANOVA y acote el P valor del test F .

Fuente	gl	SC	CM	F
Entre tratamientos	—	5,4336	—	—
Dentro de tratamientos	—	—	—	—
Total	229	73,9945	—	—

- Si $\alpha = 0,01$, ¿rechazaría la hipótesis nula de medias poblacionales iguales? ¿Por qué o por qué no?

***11.S.13** En referencia al Ejercicio 11.S.2, defina contrastes para medir cada uno de los efectos especificados, y calcule el valor de cada contraste.

- El efecto del probucol en ausencia de multivitaminas.
- El efecto del probucol en presencia de multivitaminas.
- La interacción entre el probucol y las multivitaminas.

***11.S.14** En referencia al Ejercicio 11.S.12, construya un intervalo de confianza del 95 % ($\alpha_{cw} = 0,05$) para el efecto del probucol en la ausencia de multivitaminas. Es decir, construya el intervalo de confianza del 95 % para $\mu_{\text{probuco}} - \mu_{\text{placebo}}$.

***11.S.15** En referencia al Ejercicio 11.S.2, suponiendo que se calculan todas las posibles comparaciones de las medias de los grupos, utilice el método de Bonferroni para construir un intervalo de confianza del 95 % para el efecto del probucol en ausencia de multivitaminas. Es decir, construya un intervalo de confianza del 95 % ajustado por Bonferroni ($\alpha_{ew} = 0,05$) para $\mu_{\text{probuco}} - \mu_{\text{placebo}}$.

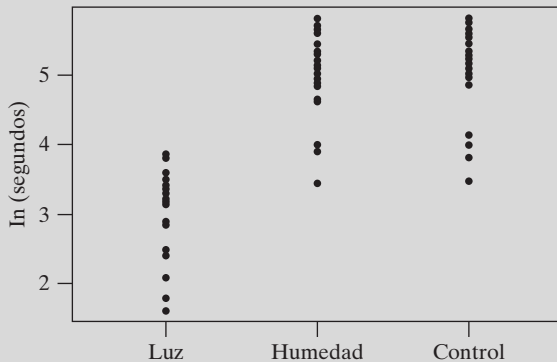
***11.S.16** Tres estudiantes universitarios recogieron varias cochinillas de una pila de leña y las utilizaron en un experimento en el que midieron el tiempo, en segundos, que empleaba una cochinilla en moverse 6 pulgadas en el interior de un aparato que habían creado. Había tres grupos de cochinillas: un grupo se expuso a una luz fuerte, para otro grupo el estímulo fue la humedad, y un tercer grupo sirvió como control. Los datos se muestran en la tabla siguiente³⁶.

	Luz	Humedad	Control
	23	170	229
	12	182	126
	29	286	140
	12	103	260
	5	330	330
	47	55	310
	18	49	45
	30	31	248
	8	132	280

(Continúa)

	Luz	Humedad	Control
	45	150	140
	36	165	160
	27	206	192
	29	200	159
	33	270	62
	24	298	180
	17	100	32
	11	162	54
	25	126	149
	6	229	201
	34	140	173
Media	23,6	169,2	173,5
DT	12,3	83,5	86,0
n	20	20	20

Las DT muestran claramente que la variabilidad no es constante entre los grupos, por lo que es necesaria una transformación. Tomando el logaritmo natural de cada observación resulta el siguiente diagrama de puntos y estadísticos resumen.



	Luz	Humedad	Control
Media	2,99	4,98	4,99
DT	0,65	0,62	0,66

Para los datos transformados, los valores del ANOVA son $SC(entre) = 53,1103$ y $SC(intra) = 23,5669$.

- Plantee con símbolos la hipótesis nula.
- Construya la tabla del ANOVA y contraste la hipótesis nula. Utilice $\alpha = 0,05$.
- Calcule la desviación típica conjunta, $s_{conjunta}$.

***11.S.17** A menudo los alpinistas experimentan diversos síntomas cuando alcanzan altitudes grandes durante sus es-

caladas. Unos investigadores estudiaron los efectos de la exposición a gran altitud en tejidos de músculos esqueléticos humanos. Realizaron un experimento factorial 2×2 en el que los sujetos se entrenaron con bicicleta durante seis semanas. El primer factor fue si los sujetos fueron entrenados bajo condiciones hipóxicas (correspondientes a una altitud de 3.850 m) o bajo condiciones normales. El segundo factor fue si los sujetos se entrenaron con un alto nivel de gasto energético o con un nivel bajo (25 % menos que los de alto nivel). Había 7 u 8 sujetos en cada combinación de niveles de los factores. La tabla que acompaña a este ejercicio muestra los resultados de la variable de respuesta «cambio porcentual en el factor de crecimiento vascular endotelial mRNA»³⁷.

Energía	Hipóxico		Normal	
	Nivel bajo	Nivel alto	Nivel bajo	Nivel alto
Media	117,7	173,2	95,1	114,6
Número de pacientes	7	7	8	8

Prepare un gráfico de interacción (como la Figura 11.7.3).

***11.S.18** Considere los datos del Ejercicio 11.S.17.

- Complete la siguiente tabla de ANOVA.

Fuente	gl	SC	CM	Razón F
Entre hipoxia y normal	1	12.126,5	—	—
Entre nivel de energía	1	10.035,7	—	—
Interacción	1	—	—	—
Dentro de grupos	26	56.076,0	—	—
Total	29	80.738,7	—	—

- Realice un test para las interacciones. Utilice $\alpha = 0,05$.
- Basándose en sus conclusiones del apartado (b), ¿es sensato examinar los efectos principales de la condición y del nivel de energía?
- Contraste la hipótesis nula de que el nivel de energía no tiene efecto en la respuesta. Utilice $\alpha = 0,05$.
- Contraste la hipótesis nula de que el efecto sobre la respuesta con entrenamiento hipóxico es el mismo que el efecto sobre la respuesta del entrenamiento normal. Utilice $\alpha = 0,05$.

***11.S.19** En un estudio para examinar la utilidad de emplear gas amoníaco para sanear el pienso de los animales, unos investigadores inocularon ensilajes de maíz con una cepa de *Salmonella*. Seguidamente, se expusieron dos placas de Petri con 5 g de comida contaminada a gas amoníaco anhidro concentrado y otras dos placas de Petri con 5 g de comida contaminada no fueron tratadas con el gas. Este experimento se repitió dos veces, con un total de tres pruebas,

ya que solo se podían situar dos placas de Petri a la vez en la cámara de gas presurizado. 24 horas después de la inoculación y el tratamiento con el gas, se contó el número de colonias de bacterias (unidades de formación de colonias o ufc) en cada placa. Debido a que los datos estaban altamente sesgados, se analizó el $\log(\text{ufc})^{38}$.

Fuente	gl	SC	CM	Razón F
Entre tratamientos	1	1,141	1,141	7,107
Entre pruebas	2	3,611	—	—
Dentro de los grupos	8	—	—	—
Total	11	6,036		

- Identifique los bloques, los tratamientos y las variables de respuesta en este problema.
- Complete la siguiente tabla de ANOVA para este análisis por bloques.
- Utilizando la tabla completa del apartado (b), ¿hay evidencia de que el tratamiento con gas amoníaco afecta al nivel de contaminación (es decir a la media del logaritmo de las ufc)? Utilice $\alpha = 0,05$.
- ¿Permiten el análisis y la información anterior inferir que el amoníaco reduce la contaminación? Si no es así, ¿qué otra información sería necesaria para realizar esa afirmación?

Notas

- Martinez, J. (1998). Organic practices for the cultivation of sweet corn. Manuscrito no publicado, Oberlin College. A efectos pedagógicos, los datos presentados aquí son una muestra aleatoria de un estudio mayor. El método utilizado fue *Steinernema carpocapsae*, la bacteria fue *Bacillus thuringiensis*, y la avispa fue *Trichogramma pretiosum*.
- Shields, D. R. (1981). The influence of niacin supplementation on growing ruminants and *in vivo* and *in vitro* rumen parameters. Ph. D. thesis, Purdue University. Adaptado de datos iniciales proporcionados por cortesía del autor y de D. K. Colby.
- Adaptado de Potkin, S. G., Cannon, H. E., Murphy, D. L. y Wyatt, R. J. (1978). Are paranoid schizophrenics biologically different from other schizophrenics? *New England Journal of Medicine* **298**, 61-66. Reimpreso con permiso. Los cálculos se basan en los datos del Ejemplo 1.1.4 de este libro, que son una reconstrucción aproximada a partir de los histogramas de información resumen dados por Potkin *et al.*
- Adaptado de Keller, S. E., Weiss, J. M., Schleifer, S. J., Miller, N. E. y Stein, M. (1981). Suppression of immunity by stress: Effect of a graded series of stressors on lymphocyte stimulation in the rat. *Science* **213**, 1397-1400. Copyright 1981 de la AAAS. Las SD y los SC se estimaron de los ET dados por Keller *et al.*
- Lobstein, D. D. (1983). A multivariate study of exercise training effects on beta-endorphin and emotionality in psychologically normal, medically healthy men. Ph.D. thesis, Purdue University. Datos iniciales cortesía del autor.
- Hayden, F. G., Osterhaus, A. D., Treanor, J. J., Fleming, D. M., Aoki, F. Y., Nicholson, K. G., Bohnen, A. M., Hirst, H. M., Keene, O. y Wightman, K. (1997). Efficacy and safety of the neu-raminidase inhibitor zanamivir in the treatment of influenzavirus infections. *New England Journal of Medicine* **337**, 874-880. Las sumas de cuadrados se han calculado a partir de las medias y DT dadas en el artículo.
- Person, A. (1999). Daffodil stem lengths. Manuscrito no publicado, Oberlin College. El conjunto de datos completo es algo mayor que el presentado aquí.
- Kotler, D. (2000). A comparison of aerobics and modern dance training on health-related fitness in college women. Manuscrito no publicado, Oberlin College.
- Datos no publicados cortesía de H. W. Ohm.
- Cameron, E. y Pauling, L. (1978). Supplemental ascorbate in the supportive treatment of cancer: Re-evaluation of prolongation of survival times in terminal human cancer. *Proceedings of the National Academy of Science USA* **75**, 4538-4542.
- Neumann, A., Richards, A.-L. y Randa, J. (2001). Effects of acid rain on alfalfa plants. Manuscrito no publicado, Oberlin College. El grupo de ácido bajo recibió 1,5 M HCL así como dos cuentagotas llenos de agua cada día. En el grupo del ácido se usó 3,0 M HCL. El grupo de control solo recibió agua. Los datos originales se modificaron ligeramente por motivos pedagógicos.
- Este es el diseño descrito en los siguientes artículos. Rosenzweig, M. R., Bennett, E. L. y Diamond, M. C. (1972). Brain changes in response to experience. *Scientific American* **226**, n.º 2, 22-29. Bennett, E. L., Diamond, M. C., Krech, D. y Rosenzweig, M. R. (1964). Chemical and anatomical plasticity of brain. *Science* **146**, 610-619.
- Basado en un experimento de Resh, W. y Stoughton, R. B. (1976). Topically applied antibiotics in acne vulgaris. *Archives of Dermatology* **112**, 182-184.
- Swearingen, M. L. y Holt, D. A. (1976). Using a «blank» trial as a teaching tool. *Journal of Agronomic Education* **5**, 3-8. Reimpreso con permiso de la American Society of Agronomy, Inc via Copyright Clearance Center. De hecho, para demostrar la variabilidad de los rendimientos de las parcelas, los investigadores plantaron la misma variedad de cebada en las 16 parcelas.

15. Datos adaptados de McIntosh, A. R. y Townsend, C. R. (1996). Interactions between fish, grazing invertebrates and algae in a New Zealand stream: A trophic cascade mediated by fish-induced changes to grazer behavior. *Oecologia* **108**, 174-181.
16. Pappas, T. y Mitchell, C. A. (1985). Effects of seis-mic stress on the vegetative growth of *Glycine max* (L.) Merr. cv. Wells II. *Plant, Cell and Environment* **8**, 143-148. Datos iniciales cortesía de los autores. El experimento original incluyó más de cuatro tratamientos. Reimpreso con permiso de Plant, Cell y Environment.
17. Garcia-Nebot, M., Alegria, A., Barbera, R., Clemente, G. y Romero, F. (2010). Addition of milk or caseinophosphopeptides to fruit beverages to improve iron bioavailability. *Food Chemistry* **119**, 141-148.
18. Kiesecker, J. M., Blaustein, A. R. y Belden, L. K. (2001). Complex causes of amphibian population declines. *Nature* **410**, 681-684. Las medias muestrales y la desviación típica se obtuvieron de la Figura 2a del artículo.
19. Tripepi, R. R. y Mitchell, C. A. (1984). Metabolic response of river birch and European birch roots to hypoxia. *Plant Physiology* **76**, 31-35. Datos iniciales cortesía de los autores.
20. Adaptado de Veterans Administration Cooperative Study Group on Antihypertensive Agents (1979). Comparative effects of ticrynafen and hydrochlorothiazide in the treatment of hyper-tension. *New England Journal of Medicine* **301**, 293-297. Reimpreso con permiso. El valor de $s_{conjunta}$ se calculó a partir de los ET dados en el artículo. Copyright © 1979 Massachusetts Medical Society. Todos los derechos reservados.
21. Knight, S. L. y Mitchell, C. A. (1983). Enhancement of lettuce yield by manipulation of light and nitrogen nutrition. *Journal of the American Society for Horticultural Science* **108**, 750-754. Cálculos basados en los datos iniciales proporcionados por los autores.
22. Datos ficticios pero realistas, adaptados de O'Brien, R. J. y Drizd, T. A. (1981). Basic data on spirometry in adults 25-74 years of age: United States, 1971-75. *U.S. National Center for Health Statistics, Vital and Health Statistics*, Serie 11, n.º 222. Washington, D.C.: U.S. Department of Health and Human Services.
23. U.S. Bureau of the Census. La distribución de edades de 2008 es de www.census.gov/compendia/satab/cats/population/estimates_and_projections_by_age_sex_raceethnicity.html Los porcentajes se han redondeado para que sumen 1.
24. Chrisman, C. L. y Baumgartner, A. P. (1980). Micronuclei in bone-marrow cells of mice subjected to hyperthermia. *Mutation Research* **77**, 95-97. El experimento original incluía seis tratamientos.
25. Baird, J. T. y Quinlivan, L. G. (1973). Parity and hypertension. *U.S. National Center for Health Statistics, Vital and Health Statistics*, Serie 11, n.º 38. Washington, D.C.: U.S. Department of Health, Education and Welfare.
26. U.S. Bureau of the Census (1997). *Statistical Abstract of the United States*, 1997, 117ª ed. Washington, D.C.: U.S. Government Printing Office.
27. Adaptado de Witelson, S. F. (1985). The brain connection: The corpus callosum is larger in left-handers. *Science* **229**, 665-668. Copyright 1985 de la AAAS. La DT y los CM (intra) se han calculado a partir de los errores típicos dados por Witelson. Reimpreso con permiso de la AAAS.
28. Adaptado de Booth, D. M. y Heck, K. L. (2009). Effects of the American oyster *Crassostrea virginica* on growth rates of the seagrass *Halodule wrightii*. *Marine Ecology Progress Series* **389**, 117-126. El artículo tiene el conjunto completo de datos; presentamos solo parte de ellos por motivos pedagógicos.
29. Latimer, J. (1985). Adaptado de datos no publicados proporcionados por el investigador.
30. Adaptado de Morris, J. G., Cripe, W. S., Chapman, H. L., Jr., Walker, D. F., Armstrong, J. B., Alexander, J. D., Jr., Miranda, R., Sanchez, A., Jr., Sanchez, B., Blair-West, J. R. y Denton, D. A. (1984). Selenium deficiency in cattle associated with Heinz bodies and anemia. *Science* **223**, 491-492. Copyright 1984 de la AAAS. La SC (intra) es ficticia pero se ajusta a los errores típicos dados por Morrisetal.
31. Datos ficticios pero realistas. Adaptados de Mizutani, T. y Mitsuoka, T. (1979). Effect of intestinal bacteria on incidence of liver tumors in gnotobiotic C3H/He male mice. *Journal of the National Cancer Institute* **63**, 1365-1370.
32. Becker, W. A. (1961). Comparing entries in random sample tests. *Poultry Science* **40**, 1507-1514.
33. Adaptado de Rosner, B. (1982). Statistical methods in ophthalmology: An adjustment for the intraclass correlation between eyes. *Biometrics* **38**, 105-114. Reprinted with permission from The International Biometric Society. El estudio médico se explica en Berson, E. L., Rosner, B. y Simonoff, E. (1980). An outpatient population of retinitis pigmentosa and their normal relatives: Risk factors for genetic typing and detection derived from their ocular examinations. *American Journal of Ophthalmology* **89**, 763-775. Las medias y las sumas de cuadrados se han estimado de los datos dados por Rosner, tras estimar los valores que faltaban de dos pacientes en los que solo existía la medida de un ojo.
34. Heggstad, H. E. y Bennett, J. H. (1981). Photochemical oxidants potentiate yield losses in snap beans attributable to sulfur dioxide. *Science* **213**, 1008-1010. Copyright 1981 de la AAAS. Datos iniciales cortesía de H. E. Heggstad.
35. Tardif, J., Cote, G., Lesperance, J., Bourassa, M., Lambert, J., Doucet, S., Bilodeau, L., Nattel, S. y DeGuise, P. (1997). Probuco and multivitamins in the prevention of

restenosis after coronary angioplasty. *New England Journal of Medicine* **337**, 365-372.

36. Walker, P., Osredkar, M. y Bilancini, S. (1999). The effect of stimuli on pillbug movement. Manuscrito no publicado, Oberlin College.

37. Hoppeler, H. y Vogt, M. (2001). Muscle tissue adaptations to hypoxia. *The Journal of Experimental Biology* **204**, 3133-3139.

38. Adaptado de Tajkarimi, M., Riemann, H., Hajmeer, M., Gomez, E., Razavilar, V. y Cliver, D. (2008). Ammonia disinfection of animal feeds-laboratory study. *International Journal of Food Microbiology* **122**, 23-28.

REGRESIÓN LINEAL Y CORRELACIÓN

12

Objetivos

En este capítulo estudiaremos la correlación y la regresión:

- demostraremos cómo se calcula e interpreta el coeficiente de correlación;
- mostraremos cómo se ajustan a los datos los modelos de regresión basados en mínimos cuadrados;
- examinaremos la relación entre la recta de regresión, la correlación muestral y la predicción de medias;
- mostraremos cómo contrastar si una relación de regresión es estadísticamente significativa;
- extendaremos las ideas a la regresión múltiple, al análisis de covarianza y a la regresión logística.

12.1 Introducción

En este capítulo presentaremos algunos métodos para analizar la relación entre dos variables cuantitativas, X e Y . La **regresión lineal** y el **análisis de correlación** son técnicas basadas en ajustar una recta a los datos.

Ejemplos

Los datos para la regresión y el análisis de correlación consisten en parejas de observaciones (X, Y) . He aquí dos ejemplos.

Ejemplo 12.1.1 Anfetamina y consumo de alimentos

La anfetamina es una sustancia que suprime el apetito. En un estudio sobre este efecto, una farmacóloga asignó aleatoriamente 24 ratas a tres grupos de tratamiento, de forma que recibirían una inyección de anfetamina con dos posibles dosificaciones, o bien una inyección de solución salina. Midió la cantidad de alimento consumido por cada animal en las tres horas siguientes a la administración de la inyección. Los resultados (g de alimento consumido por kilogramos de peso corporal) se muestran en la Tabla 12.1.1¹.

La Figura 12.1.1 muestra un diagrama de dispersión de

$Y =$ Consumo de alimento

frente a

$X =$ Dosis de anfetamina

El diagrama de dispersión sugiere una relación definida dosis-respuesta, en la que los valores mayores de X tienden a estar asociados con valores menores de Y^* . ■

Tabla 12.1.1 Consumo de alimento (Y) de ratas (g/kg)			
	X = dosis de anfetamina (mg/kg)		
	0	2,5	5,0
	112,6	73,3	38,5
	102,1	84,8	81,3
	90,2	67,3	57,1
	81,5	55,3	62,3
	105,6	80,7	51,5
	93,0	90,0	48,3
	106,6	75,5	42,7
	108,3	77,1	57,9
Media	100,0	75,5	55,0
DT	10,7	10,7	13,3
Número de animales	8	8	8

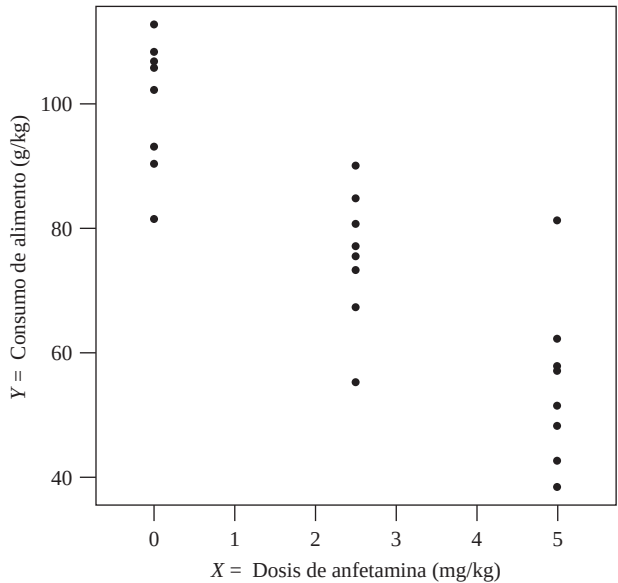


Figura 12.1.1 Diagrama de dispersión del consumo de alimento frente a la dosis de anfetamina

Ejemplo 12.1.2 Arsénico en arroz

Los contaminantes ambientales pueden entrar en la cadena alimentaria por filtración en el suelo donde se cultivan los alimentos. Se tiene la hipótesis de que el silicio natural en las plantas de arroz puede inhibir la absorción de

* En muchas relaciones dosis-respuesta, la respuesta depende linealmente del $\log(\text{dosis})$ en vez de depender de la propia dosis. Hemos elegido una porción lineal de la curva dosis-respuesta para simplificar la exposición.

algunos contaminantes. En un estudio para investigar compuestos que podrían mitigar la absorción de arsénico por el arroz, unos investigadores muestrearon 32 plantas de arroz y midieron en cada planta la concentración de arsénico en el arroz limpio ($\mu\text{g/kg}$ de arroz), así como la concentración de silicio en la paja (g/kg de paja)². La Figura 12.1.2 muestra un diagrama de dispersión de

$Y = \text{Concentración de arsénico en el arroz}$

frente a

$X = \text{Concentración de silicio en la paja}$

El diagrama de dispersión sugiere que concentraciones mayores de silicio en la paja (X) tienden a estar asociadas con menores concentraciones de arsénico (Y). ■

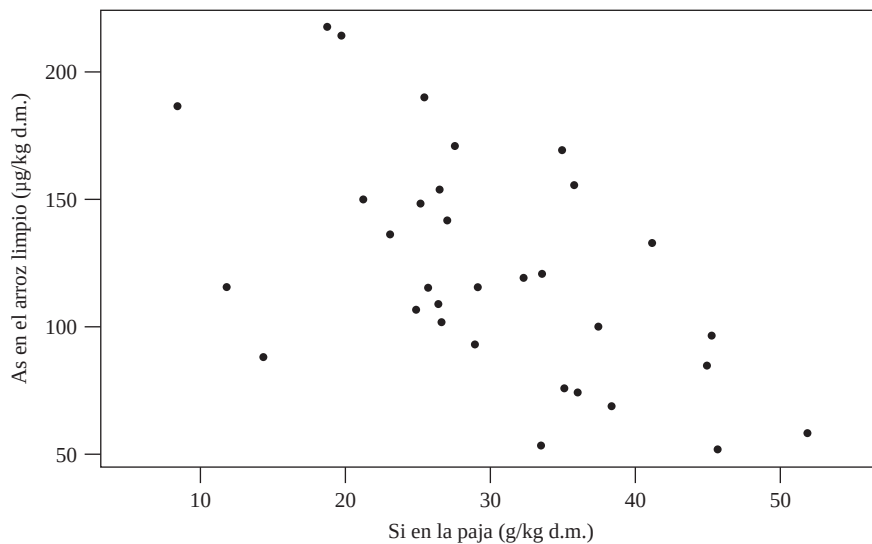


Figura 12.1.2 Diagrama de dispersión de la concentración de arsénico en el arroz frente a la concentración de silicio en la paja

12.2 El coeficiente de correlación

Supongamos que tenemos una muestra de n parejas en donde cada pareja representa las medidas de dos variables, X e Y . Si un diagrama de dispersión de Y frente a X muestra una tendencia general lineal, entonces resulta natural intentar describir la fuerza de la asociación lineal. En esta sección aprenderemos a medir la fuerza de la asociación lineal utilizando el **coeficiente de correlación**. El ejemplo siguiente ilustra el tipo de situación que deseamos considerar.

Ejemplo 12.2.1 Longitud y peso de serpientes

En un estudio de la población salvaje de la serpiente *Vipera bertis*, unos investigadores capturaron y midieron nueve hembras adultas³. Las longitudes de sus cuerpos y sus pesos se muestran en la Tabla 12.2.1 y se representa en forma de diagrama de dispersión en la Figura 12.2.1. El número de observaciones es $n = 9$. ■

El diagrama de dispersión que se muestra en la Figura 12.2.1 indica una clara tendencia ascendente. Se dice que el peso muestra una **asociación positiva** con la longitud, lo que indica que longitudes mayores están asociadas con pesos mayores. Por tanto, la serpientes que son más largas que la longitud promedio de $\bar{x} = 63$ tienden a pesar más que el peso promedio de $\bar{y} = 152$. La línea superpuesta en la gráfica se denomina **recta de mínimos cuadrados** o **recta de regresión ajustada** de Y con respecto a X . Aprenderemos cómo calcular e interpretar la recta de regresión en la Sección 12.3.

Medida de la fuerza de una asociación lineal

¿Qué fuerza tiene la relación lineal entre la longitud y el peso de las serpientes? ¿Están los puntos de datos situados muy cerca de la recta de regresión o la dispersión es amplia? Para responder a estas preguntas, calcularemos el **coeficiente de correlación**, una medida numérica invariante con la escala de la fuerza de la relación lineal entre dos variables cuantitativas. El que sea invariante con la escala significa que los cambios en las escalas de medida no afectan al coeficiente de correlación. Es decir, la correlación entre la longitud y el peso será la misma tanto si se miden en centímetros y gramos como si se miden en pulgadas y libras. Para entender cómo funciona el coeficiente de correlación, consideremos de nuevo el ejemplo de la longitud y el peso de las serpientes. En vez de representar los datos originales, en la Figura 12.2.2 se representan los datos estandarizados (puntuaciones z) que se muestran en la Tabla 12.2.2. Nótese que esta gráfica tiene un aspecto idéntico al de nuestra gráfica original excepto porque ahora nuestras escalas son adimensionales.

Tabla 12.2.1		
	Longitud X (cm)	Peso Y (g)
	60	136
	69	198
	66	194
	64	140
	54	93
	67	172
	59	116
	65	174
	63	145
Media	63	152
DT	4,6	35,3

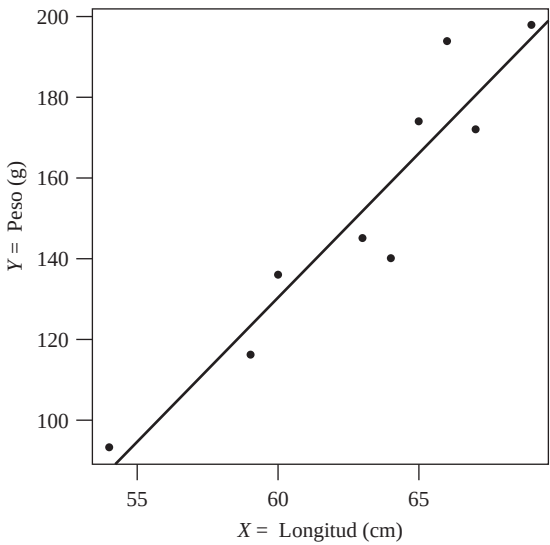


Figura 12.2.1 Longitud corporal y peso de nueve serpientes con una recta de regresión ajustada

Dividiendo la gráfica en cuadrantes basándose en el signo de la puntuación estandarizada, vemos que la mayoría de esos puntos están en los cuadrantes superior derecho e inferior izquierdo. Los puntos que están en esos cuadrantes tendrán puntuaciones estandarizadas cuyos *productos* serán positivos. Asimismo, los puntos que están en los cuadrantes superior izquierdo e inferior derecho tendrán puntuaciones estandarizadas cuyos productos serán negativos. Calculando la suma de estos productos se obtiene una medida numérica de dónde están los puntos (es decir, qué cuadrantes son dominantes). En nuestro caso, como hay una asociación positiva entre la longitud y el peso, la mayoría de los puntos están en los cuadrantes de producto positivo. Por tanto, la suma de los productos de las puntuaciones estandarizadas es positiva. Si estuviera presente una relación negativa, la mayoría de los puntos estarían en los cuadrantes negativos y la suma sería negativa. Y si no hubiera relación *lineal*, los puntos podrían estar en cualquiera de los cuatro cuadrantes de forma que los productos positivos y negativos se equilibrarían y su suma sería cero.

El coeficiente de correlación se basa en esta suma. Se calcula como el promedio de los productos de las puntuaciones estandarizadas (utilizando $n - 1$ en lugar de n para calcular el promedio)*.

* Sustituyendo s_x por $\sqrt{\sum_{i=1}^n (x - \bar{x})^2 / n - 1}$ y s_y por $\sqrt{\sum_{i=1}^n (y - \bar{y})^2 / n - 1}$, la ecuación del coeficiente de correlación se puede expresar como
$$r = \frac{\sum_{i=1}^n (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^n (x - \bar{x})^2 \sum_{i=1}^n (y - \bar{y})^2}}.$$

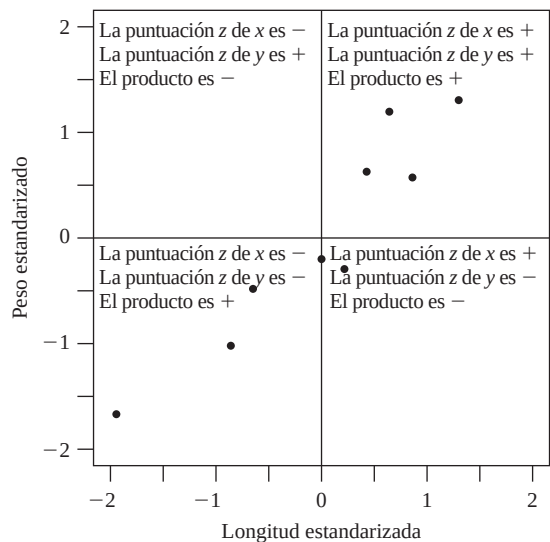


Figura 12.2.2 Diagrama de dispersión del peso estandarizado frente a la longitud estandarizada

Tabla 12.2.2 Pesos y longitudes de serpientes estandarizados y sus productos					
	Peso	Longitud	Peso estandarizado	Longitud estandarizada	Producto de valores estandarizados
	X	Y	$z_x = \frac{x - \bar{x}}{s_x}$	$z_y = \frac{y - \bar{y}}{s_y}$	$z_x z_y$
	60	136	-0,65...	-0,45...	0,29...
	69	198	1,29...	1,30...	1,68...
	66	194	0,65...	1,19...	0,77...
	64	140	0,22...	-0,34...	-0,07...
	54	93	-1,94...	-1,67...	3,24...
	67	172	0,86...	0,57...	0,49...
	59	116	-0,86...	-1,02...	0,88...
	65	174	0,43...	0,62...	0,27...
	63	145	0,00...	-0,20...	0,00...
Suma	567	1.368	0,00	0,00	7,5494
Media	63,000	152,000	0,00	0,00	
DT	4,637	35,338	1,00	1,00	

Los valores de la tabla se han truncado para facilitar su lectura. Debido a que los valores resumen se utilizarán en cálculos posteriores, incluye más dígitos que los que normalmente se utilizarían en un informe siguiendo nuestros convenios sobre el redondeo.

El coeficiente de correlación, r

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x - \bar{x}}{s_x} \right) \left(\frac{y - \bar{y}}{s_y} \right)$$

Resulta claro que X e Y entran en la fórmula de r simétricamente. Por tanto, si intercambiamos las etiquetas X e Y de nuestras variables, r no cambiaría. De hecho, esta es una de las ventajas del coeficiente de correlación como resumen estadístico. Al interpretar r , no es necesario saber (o decidir) qué variable se etiqueta como X y qué variable se etiqueta como Y .

Interpretación del coeficiente de correlación

Matemáticamente, el coeficiente de correlación no tiene unidades y siempre está entre -1 y 1 . El signo de la correlación indica el signo de la relación y coincide con el signo de la pendiente de la recta de regresión: positivo (creciente) o negativo (decreciente). Cuanto más cerca está la correlación de -1 o de 1 , más fuerte es la relación lineal entre X e Y . Una correlación igual a -1 o a 1 indica una relación perfectamente lineal entre las dos variables (un diagrama de dispersión de esos datos mostraría que están exactamente sobre una línea recta). Es interesante indicar que una correlación de cero no significa necesariamente que no haya relación entre X e Y , sino que solo significa que no hay relación lineal entre X e Y . El cálculo anterior de la correlación indica que la suma de los productos de los valores estandarizados era cero siempre que los productos positivos y negativos se compensen. Esto puede suceder de muchas formas. La Figura 12.2.3 muestra varios ejemplos con diversos valores del coeficiente de correlación.

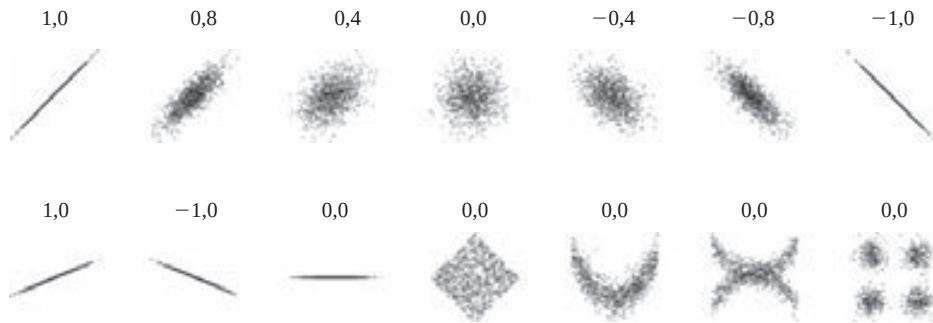


Figura 12.2.3 Diagramas de dispersión de datos con varios valores de la correlación muestral

Ejemplo 12.2.2 Longitud y peso de serpientes

En la Tabla 12.2.2 mostramos que para los datos de las serpientes la suma de los productos de las puntuaciones estandarizadas es 7,5494. Por tanto, el coeficiente de correlación de las longitudes y pesos de nuestra muestra de nueve serpientes es aproximadamente 0,94.

$$r = \frac{1}{9 - 1} \times 7,5494 \approx 0,94$$

En este ejemplo, podemos denominar también al valor de 0,94 **correlación muestral**, ya que las longitudes y los pesos de las nueve serpientes forman una muestra de una población mayor. La correlación muestral es un estimador de la **correlación poblacional** (que frecuentemente se denomina con la letra griega «rho», ρ) en este caso el coeficiente de correlación de la población completa de serpientes *Vipera bertis* hembra. Para considerar al coeficiente de correlación r como un estimador de un parámetro poblacional, debe ser razonable suponer que tanto los valores de X como los de Y se seleccionan aleatoriamente, como en el siguiente **modelo de muestreo aleatorio bivariado**:

Modelo de muestreo aleatorio bivariado:

Consideramos que cada pareja (x_i, y_i) se ha muestreado aleatoriamente de una población de parejas (x, y) .

En el modelo de muestreo aleatorio bivariado, las X observadas se consideran una muestra aleatoria y las Y observadas se consideran también una muestra aleatoria, de forma que los estadísticos marginales $\bar{x}$, $\bar{y}$, s_x y s_y son estimadores de los correspondientes valores poblacionales μ_x , μ_y , σ_x y σ_y .

Para muchos investigadores el modelo de muestreo aleatorio es razonable, pero el supuesto adicional de un modelo de muestreo aleatorio bivariado no lo es. Este es en general el caso cuando los valores de X los fija el investigador como en el Ejemplo 12.1.1, en el que los investigadores asignaron ratas a tres dosis de anfetamina. Este tipo de modelo de muestreo se denomina modelo de submuestreo aleatorio y se define en la Sección 12.4. En estos casos el coeficiente de correlación muestral no es un estimador adecuado de la correlación poblacional.

Inferencia con respecto a la correlación

Hemos presentado la forma en la que el coeficiente de correlación describe un conjunto de datos dentro del modelo de muestreo aleatorio bivariado. Veremos ahora la inferencia estadística basada en r con este modelo. Consideremos el siguiente ejemplo.

Contraste de la hipótesis $H_0: \rho = 0$

En algunas investigaciones no se conoce de antemano si existe una relación entre X e Y . En ese caso puede ser relevante considerar la posibilidad de que cualquier tendencia aparente en los datos es ilusoria y refleja solo la variabilidad del muestreo. En esta situación resulta natural formular la hipótesis nula

H_0 : X e Y están incorrelacionados en la población

o, de forma equivalente,

H_0 : no existe relación lineal entre X e Y

El test t de H_0 se basa en el estadístico de contraste

$$t_s = r \sqrt{\frac{n-2}{1-r^2}}$$

Los valores críticos se obtienen de la distribución t de Student con

$$gl = n - 2$$

El ejemplo siguiente ilustra la aplicación de este test.

Ejemplo 12.2.3 Presión sanguínea y calcio en las plaquetas

Se sospecha que el calcio en las plaquetas de la sangre puede estar relacionado con la presión sanguínea. Como parte de un estudio sobre esta relación, unos investigadores reclutaron a 38 sujetos con presión sanguínea normal (es decir, no anormalmente elevada)⁴. Se tomaron dos medidas de cada sujeto: presión (promedio de las medidas sistólica y diastólica) y concentración de calcio en las plaquetas de la sangre. Los datos se muestran en la Figura 12.2.4. La muestra es de tamaño $n = 38$ y la correlación muestral es $r = 0,5832$.

¿Hay evidencia de que la presión sanguínea y el calcio en las plaquetas están relacionados linealmente? Contrastaremos la hipótesis nula

$$H_0: \rho = 0$$

contra la alternativa no direccional

$$H_A: \rho \neq 0$$

Esas hipótesis son traducciones de las hipótesis verbales:

H_0 : el calcio en las plaquetas no está relacionado linealmente con la presión sanguínea.

H_A : el calcio en las plaquetas está relacionado linealmente con la presión sanguínea.

Escojamos $\alpha = 0,05$. El estadístico de contraste es

$$t_s = 0,5832 \sqrt{\frac{38-2}{1-0,5832^2}} = 4,308$$

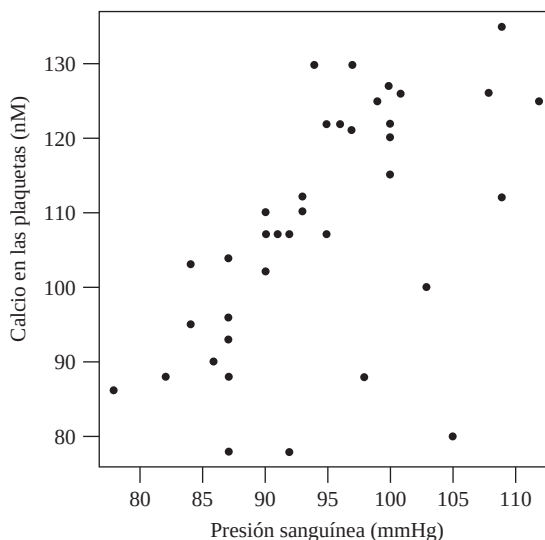


Figura 12.2.4 Presión sanguínea y calcio en las plaquetas de 38 personas con presión sanguínea normal

Utilizando la Tabla 4 con $gl = n - 2 = 36 \approx 40$, obtenemos $t_{40, 0,0005} = 3,551$. Por tanto, tenemos que P valor $< 0,0005 \times 2 = 0,001$ (ya que es no direccional) y rechazamos H_0 . Los datos proporcionan evidencia fuerte a favor de que el calcio en las plaquetas está relacionado linealmente con la presión sanguínea ($t_s = 4,308$, $gl = 36$, P valor $< 0,001$). ■

¿Por qué $n - 2$? El estadístico t del contraste de hipótesis para la población anterior tiene asociados $gl = n - 2$. El origen del valor $n - 2$ es fácil de explicar. Dos puntos cualesquiera determinan una línea recta, por lo que un conjunto de datos tan pequeño como de $n = 2$ no proporciona información sobre la variabilidad inherente del diagrama de dispersión de los puntos (o, de forma equivalente, sobre la fuerza de la asociación entre X e Y). Hasta que no observamos un tercer punto no podemos empezar a estimar la fuerza de la relación. Como en nuestros contextos anteriores relacionados con las distribuciones t y las distribuciones F (Capítulos 6, 7, 8 y 11), los grados de libertad son las piezas de información proporcionadas por los datos acerca del «ruido» a partir del que el investigador desea extraer la «señal».

Intervalo de confianza para ρ (opcional)

Si el tamaño de la muestra es grande, es posible construir un intervalo de confianza del 95 % para ρ . La distribución muestral del coeficiente de correlación muestral, r , está sesgada, por lo que para construir el intervalo de confianza aplicamos sobre r la denominada transformación de Fisher:

$$z_r = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$$

donde $\ln$ es el logaritmo natural (base e). Podemos ahora construir un intervalo de confianza del 90 % para $\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$ como

$$z_r \pm 1,96 \frac{1}{\sqrt{n-3}}$$

Finalmente, podemos transformar los límites del intervalo ρ de confianza para $\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$ despejando ρ en las ecuaciones dadas por

$$\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) = z_r \pm 1,96 \frac{1}{\sqrt{n-3}}$$

Los intervalos con otros niveles de confianza se construyen análogamente. Por ejemplo, para construir un intervalo de confianza del 90 %, se sustituye 1,96 por 1,645. La construcción de un intervalo de confianza para el coeficiente de correlación se ilustra en el Ejemplo 12.2.4.

Ejemplo 12.2.4 Presión sanguínea y calcio en las plaquetas

Para los datos del Ejemplo 12.2.3, el tamaño muestral es $n = 38$ y la correlación muestral es $r = 0,5832$. La transformación de Fisher sobre r resulta

$$z_r = \frac{1}{2} \ln \left(\frac{1 + 0,5832}{1 - 0,5832} \right) = \frac{1}{2} \ln \left(\frac{1,5832}{0,4168} \right) = 0,6673$$

El intervalo de confianza del 95 % para $\frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right)$ es

$$0,6673 \pm 1,96 \frac{1}{\sqrt{38 - 3}} = 0,667 \pm 0,3313$$

es decir, (0,3360, 0,9986).

Resolviendo

$$\frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right) = 0,3360 \text{ resulta } \rho = \frac{e^{2(0,3360)} - 1}{e^{2(0,3360)} + 1} = 0,32$$

Resolviendo

$$\frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right) = 0,9986 \text{ resulta } \rho = \frac{e^{2(0,9986)} - 1}{e^{2(0,9986)} + 1} = 0,76$$

Tenemos una confianza del 95 % de que la correlación entre la presión sanguínea y el calcio de las plaquetas en la población esté entre 0,32 y 0,76. Por tanto, un intervalo de confianza del 95 % para ρ es (0,32, 0,76). ■

Correlación y causalidad

Ya hemos visto anteriormente que una asociación observada entre dos variables no indica necesariamente la existencia de una conexión causal entre las mismas. Es importante recordar esta cautela al interpretar la correlación. El siguiente ejemplo muestra que incluso variables fuertemente correlacionadas pueden no estar relacionadas causalmente.

Ejemplo 12.2.5 Reproducción de un alga

Los akinetes son estructuras reproductivas como esporas producidas por el alga verde *Pithopora oedogonia*. En un estudio del ciclo de vida del alga, unos investigadores contabilizaron los akinetes en especímenes de algas obtenidos en un lago de Indiana, en 26 ocasiones durante un periodo de 17 meses. Números bajos indican germinación de los akinetes. Los investigadores apuntaron también la temperatura del agua y el fotoperiodo (horas de luz diurna) en cada una de las 26 ocasiones. Los datos mostraron una correlación negativa bastante fuerte entre el número de akinetes y el fotoperiodo. El coeficiente de correlación fue de $r = -0,72$. Los investigadores, sin embargo, se dieron cuenta de que esta correlación observada podría no reflejar una relación causal. Los días más largos (mayor fotoperiodo) tienden a coincidir con mayores temperaturas, y los akinetes podrían estar respondiendo a la temperatura en vez de al fotoperiodo. Para resolver la cuestión, los investigadores realizaron experimentos de laboratorio en los que se varió independientemente la temperatura y el fotoperiodo. Esos experimentos demostraron que el agente⁵ causal era la temperatura, y no el fotoperiodo. ■

Como muestra el Ejemplo 12.2.5, una forma de establecer la causalidad es realizar un experimento controlado donde se varíe el posible factor causal y todos los demás factores o bien se mantengan constantes o se controlen mediante aleatorización. Cuando no es posible realizar ese experimento, planteamientos indirectos utilizando análisis estadístico pueden arrojar alguna luz sobre relaciones potencialmente causales. (Uno de esos planteamientos se ilustrará en el Ejemplo 12.8.3).

Notas de advertencia

Para describir los resultados de un contraste sobre un coeficiente de correlación, los investigadores utilizan a menudo el término *significativo*, que puede ser equívoco. Por ejemplo, una afirmación como «se advierte una correlación altamente significativa» se puede malinterpretar fácilmente. Es importante recordar que la significación estadística significa simplemente el rechazo de una hipótesis nula. No implica necesariamente un efecto grande o importante. Una correlación «significativa» puede de hecho ser muy débil. Su «significación» solo significa que no se puede explicar fácilmente como

debida al azar. A partir de la fórmula $t_s = r \sqrt{\frac{n-2}{1-r^2}}$ podemos ver que para un valor fijo de r , t_s crece cuando n crece. Por tanto, si el tamaño muestral es suficientemente grande, t_s siempre podrá ser lo suficientemente grande para decir que la correlación es «significativa», independientemente de lo pequeño que sea r . Siempre es acertado evaluar la significación práctica de cualquier resultado considerando el intervalo de confianza para el parámetro poblacional de interés.

El coeficiente de correlación es altamente sensible a los puntos extremos. Por ejemplo, la Figura 12.2.5(a) muestra un diagrama de dispersión de 25 puntos con una correlación de $r = 0,2$. Uno de los puntos se ha marcado en azul. La Figura 12.2.5(b) muestra los mismos puntos, excepto porque el punto azul se ha cambiado de sitio. El cambio de ese único punto hace que el coeficiente de correlación suba de 0,2 a 0,6. La Figura 12.2.5(c) muestra una tercera versión de los datos. En este caso $r = -0,1$. Estas tres gráficas ilustran cómo un único punto puede influir grandemente en el valor del coeficiente de correlación. Siempre es importante representar los datos antes de utilizar r (o cualquier otro estadístico) para resumirlos.

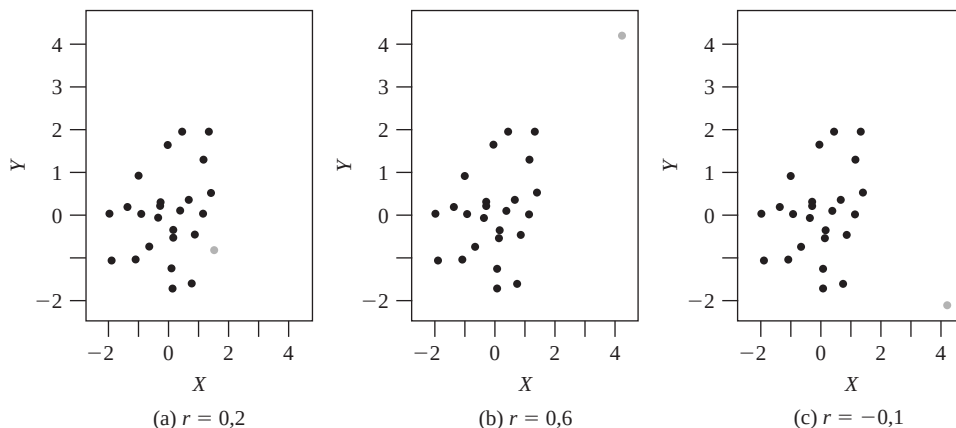
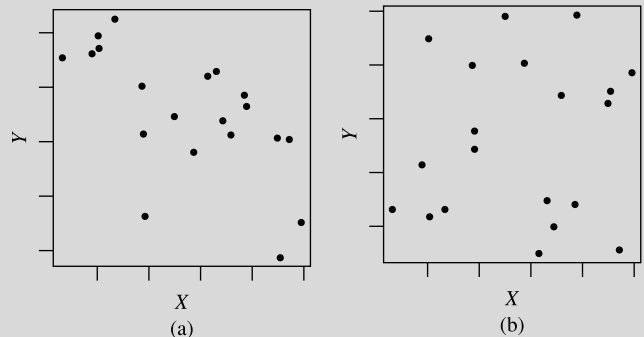
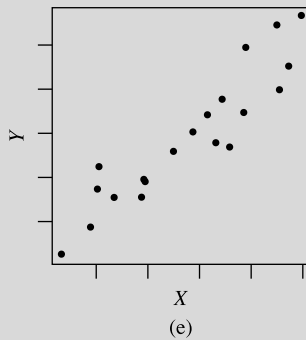
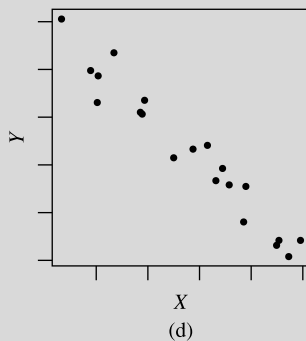
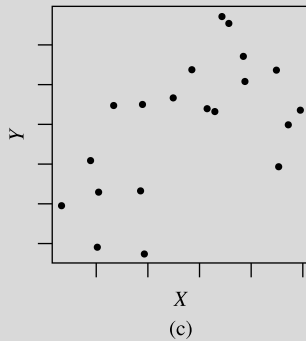


Figura 12.2.5 Efecto de *outliers* en el coeficiente de correlación muestral

Ejercicios 12.2.1-12.2.10

12.2.1 Ordene las siguientes gráficas en orden creciente de correlación (desde la más cercana a -1 hasta la más cercana a 1).





12.2.2 Considere los datos siguientes:

	X	Y
	6	6
	1	7
	3	3
	2	2
	5	14
Media	3,4	6,4
DT	2,1	4,7

- (a) Dibuje los datos. ¿Parece haber una relación entre X e Y ? ¿Es lineal o no lineal? ¿Débil o fuerte?
- (b) Calcule el coeficiente de correlación muestral entre X e Y .
- (c) ¿Hay evidencia significativa de que X e Y están correlacionadas? Realice un contraste utilizando $\alpha = 0,05$.

12.2.3 En un estudio sobre la variación natural de la química de la sangre, se obtuvieron especímenes de sangre de 284 personas sanas. Se midieron en cada espécimen las concentraciones de urea y de ácido úrico, y se encontró que la correlación entre las dos medidas era $r = 0,2291$. Contraste la hipótesis de que el coeficiente de correlación poblacional es cero contra la alternativa de que es positivo⁶. Utilice $\alpha = 0,05$.

12.2.4 Unos investigadores midieron el número de neuronas en la región CA1 del hipocampo de los cerebros de 8 personas fallecidas por causas no relacionadas con la función cerebral. Obtuvieron que esos datos estaban correlacionados negativamente con la edad⁷. El valor muestral de r fue de $-0,63$.

- (a) ¿Es este coeficiente de correlación significativamente diferente de cero? Realice un contraste con $\alpha = 0,10$.
- (b) Suponga que en el apartado (a) obtiene que la correlación difiere significativamente de cero. ¿Proporciona esto evidencia de que la edad es una causa de la pérdida de neuronas en la región CA1? Si no es así, qué se podría decir. Explique brevemente su respuesta.

12.2.5 Se escogieron aleatoriamente 20 parcelas, cada una de 10×4 metros, en un gran campo de maíz. Para cada parcela, se observaron la densidad de plantas (número de plantas en la parcela) y el peso medio de las mazorcas (g de grano por mazorca). Los resultados se muestran en la tabla⁸.

Densidad de plantas X	Peso de la mazorca Y	Densidad de plantas X	Peso de la mazorca Y
137	212	173	194
107	241	124	241
132	215	157	196
135	225	184	193
115	250	112	224
103	241	80	257
102	237	165	200
65	282	160	190
149	206	157	208
85	246	119	224

Los cálculos preliminares produjeron los siguientes resultados:

$$\begin{aligned}\bar{x} &= 128,05 & \bar{y} &= 224,10 \\ s_x &= 32,61332 & s_y &= 24,95448 \\ r &= -0,94180\end{aligned}$$

- (a) ¿Hay evidencia significativa de relación lineal entre el peso de la mazorca y la densidad de plantas? Realice un test apropiado con $\alpha = 0,05$.
- (b) ¿Es este estudio observacional o experimental?
- (c) Los granjeros están interesados en saber si la variación de la densidad de plantas puede alterar el peso de la mazorca. ¿Podrían utilizarse estos datos para responder a esta pregunta? Si no es así, ¿qué se podría decir? Explique brevemente su respuesta.

12.2.6 El ácido laetisárico es un compuesto que tiene propiedades prometedoras en el control de las enfermedades causadas por hongos en plantas de maíz. Los datos que acompañan a este ejercicio muestran los resultados de crecimiento del hongo *Phytium ultimum* en ácido laetisárico. Cada uno de los valores de crecimiento es el promedio de cuatro medidas radiales de una colonia de *P. ultimum* cultivada en una placa de Petri durante 24 horas. Había dos placas de Petri para cada concentración⁹.

- (a) ¿Hay evidencia significativa de una relación lineal entre el crecimiento de hongos y la concentración de ácido? Realice el contraste adecuado con $\alpha = 0,05$.
- (b) ¿Es este estudio observacional o experimental?
- (c) Se ha sugerido que el ácido laetisárico se podría utilizar para retrasar el crecimiento de los hongos. Podrían utilizarse estos datos para responder a esta pregunta? Si no es así, ¿qué se podría decir? Explique brevemente su respuesta.

	Concentración de ácido laetisárico X ($\mu\text{G/ml}$)	Crecimiento de hongos Y (mm)
	0	33,3
	0	31,0
	3	29,8
	3	27,8
	6	28,0
	6	29,0
	10	25,5
	10	23,8
	20	18,3
	20	15,5
	30	11,7
	30	10,0
Media	11,500	23,642
DT	10,884	7,8471
$r = -0,98754$		

12.2.7 Para investigar la dependencia del gasto de energía con la constitución corporal, unos investigadores utilizaron técnicas de pesaje bajo el agua para determinar la masa corporal libre de grasa de siete hombres. Midieron también el gasto total de energía durante 24 horas en condiciones de actividad tranquila y sedentaria. Los resultados se muestran en la tabla siguiente¹⁰. (Véase también el Ejercicio 12.5.5).

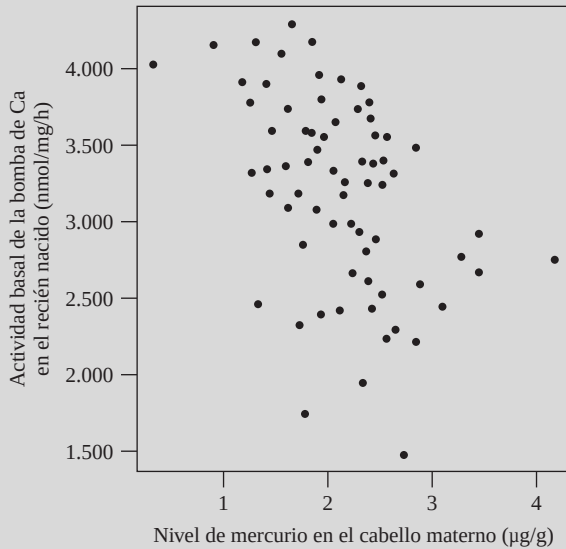
Sujeto	Masa libre de grasa X (kg)	Gasto de energía Y (kcal)
1	49,3	1.894
2	59,3	2.050
3	68,3	2.353
4	48,1	1.838
5	57,6	1.948
6	78,1	2.528
7	76,1	2.568
Media	62,400	2.168,429
DT	12,095	308,254
$r = 0,98139$		

- (a) La correlación entre el gasto de energía y la masa libre de grasa es muy grande (casi 1). Es 0,98139, pero el tamaño muestral es muy pequeño, solo de 7. ¿Hay evidencia significativa para afirmar que la correlación es diferente de cero? Realice el contraste apropiado $\alpha = 0,05$.
- (b) ¿Es este estudio observacional o experimental?
- (c) Las personas que realizan ejercicio pueden aumentar su masa libre de grasa. ¿Se podría usar este dato para afirmar que su gasto de energía también crecería? Si no es así, ¿qué se podría decir? Explique brevemente su respuesta.

12.2.8 La capacidad celular para regular la homeostasis se mide mediante la actividad basal de la bomba de Ca. La desregulación de la homeostasis del calcio puede llevar a serios efectos en el funcionamiento de las células. ¿Puede la exposición materna al mercurio medida por los depósitos de mercurio en el cabello ($\mu\text{g/g}$) afectar a la actividad basal de la bomba de Ca en los recién nacidos? El siguiente resumen de datos y gráfica provienen de un estudio en humanos con una muestra de 75 recién nacidos y sus madres¹¹.

$$\begin{aligned}\bar{x} &= 2,11183 & \bar{y} &= 3.196,8196 \\ s_x &= 0,61166 & s_y &= 611,34876 \\ r &= -0,45289\end{aligned}$$

- (a) Siempre es un buen hábito representar nuestros datos antes de analizarlos. Examinando el diagrama de dispersión, ¿parece existir una tendencia lineal en los datos? ¿Es creciente o decreciente? ¿Es débil o fuerte?



- (b) Examinando la gráfica, vemos que hay una madre con un nivel de mercurio en el cabello de aproximadamente 4,2 µg/g. Si la actividad basal de la bomba de Ca de su hijo cambiara de aproximadamente 2.800 a 2.000 nmol/mg/h, ¿aumentaría o disminuiría la correlación?
- (c) ¿Existe evidencia de que la actividad basal de la bomba de Ca de los recién nacidos disminuye con el nivel de mercurio en el cabello de la madre? Realice el contraste apropiado utilizando $\alpha = 0,05$.
- (d) En el apartado (c) debería haber obtenido que hay una evidencia fuerte a favor de una relación lineal decre-

ciente entre X e Y . Explique cómo puede ser la evidencia tan fuerte incluso aunque la gráfica de los datos muestra bastante dispersión y la correlación muestral no es cercana a -1 .

- (e) Basándose en su respuesta al apartado (c) y en el diseño de este estudio, ¿qué se podría decir con respecto a la cuestión principal de la investigación: si existe evidencia estadística de que la exposición materna al mercurio medida mediante los depósitos de mercurio en el cabello (µg/g) *afecta* a la actividad basal de la bomba de Ca de los recién nacidos (nmol/mg/h)?

12.2.9 En cada uno de los siguientes ejemplos, explique si es razonable o no tratar el coeficiente de correlación muestral, r , como un estimador del coeficiente de correlación poblacional ρ . Justifique brevemente su respuesta.

- (a) Los datos sobre la química de la sangre del Ejercicio 12.2.3.
- (b) Los datos sobre neuronas CA1 del Ejercicio 12.2.4.
- (c) Los datos sobre pesos de mazorcas del Ejercicio 12.2.5.
- (d) Los datos sobre crecimiento de hongos del Ejercicio 12.2.6.
- (e) La actividad de la bomba basal de Ca del Ejercicio 12.2.8.

12.2.10 (opcional) Para cada uno de los siguientes conjuntos de datos, calcule un intervalo de confianza del 95 % para el coeficiente de correlación poblacional.

- (a) Los datos sobre la química de la sangre del Ejercicio 12.2.3.
- (b) Los datos sobre pesos de mazorcas del Ejercicio 12.2.5.
- (c) Los datos sobre gasto de energía del Ejercicio 12.2.7.

12.3 La recta de regresión

En la Sección 12.2 aprendimos cómo el coeficiente de correlación describe la fuerza de la asociación lineal entre dos variables numéricas, X e Y . En esta sección aprenderemos cómo obtener e interpretar la recta que mejor resume esta relación lineal.

Ejemplo 12.3.1 Temperatura del océano

Consideremos un conjunto de datos en el que hay una relación lineal perfecta entre X e Y por ejemplo, temperatura medida en $X = \text{Celsius}$ e $Y = \text{Fahrenheit}$. La Figura 12.3.1 muestra 20 temperaturas oceánicas semanales (en °C y °F) en una ciudad de la costa de California, junto con una recta que describe perfectamente la relación*: $y = 32 + \frac{9}{5}x$. La Tabla 12.3.1 presenta un resumen de los datos¹².

Como X e Y miden la misma variable (temperatura), parece razonable que un espécimen de agua que tenga una DT por encima del promedio en °C ($s_x = 2,60$) tenga también una DT por encima del promedio en °F ($s_y = 2,88$). Si se combinan, estos valores pueden describir la pendiente de la recta que se ajusta exactamente a los datos

$$\frac{\text{elevación}}{\text{desplazamiento lateral}} = \frac{s_y}{s_x} = \frac{2,88}{1,60} = 1,80$$

* Esta ecuación es la fórmula de conversión de Celsius a Fahrenheit.

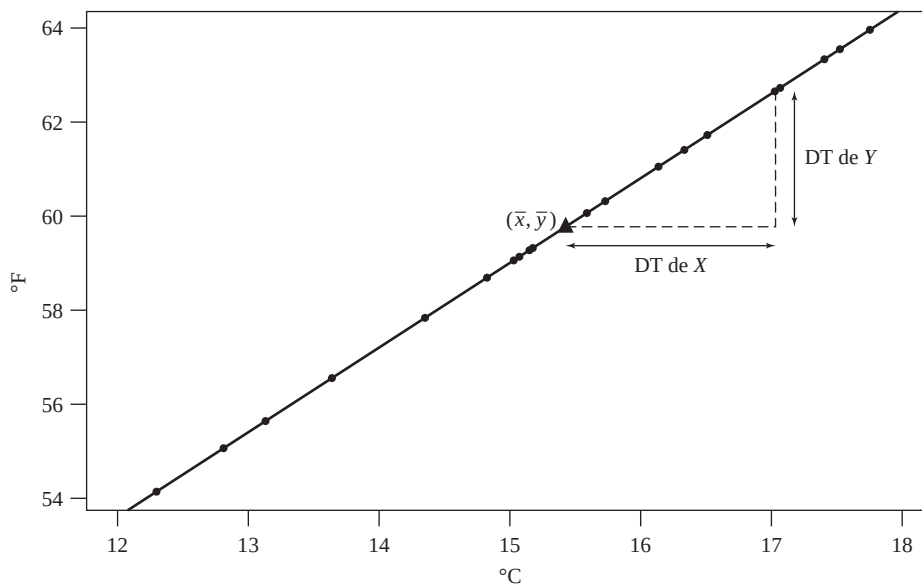


Figura 12.3.1 Diagrama de dispersión de Y = temperatura oceánica en $^{\circ}\text{F}$ frente a X = temperatura oceánica en $^{\circ}\text{C}$. El valor medio $(\bar{x}, \bar{y})$ se indica con ▲

Tabla 12.3.1 Resumen de los datos de temperatura del agua		
	X = temperatura ($^{\circ}\text{C}$)	Y = temperatura ($^{\circ}\text{F}$)
Media	15,43	59,77
DT	1,60	2,88

En este ejemplo sucede que conocemos la ecuación de la recta utilizada para convertir de Celsius a Fahrenheit. La pendiente de esta recta es $9/5 = 1,80$, el mismo valor obtenido anteriormente. ■

La recta DT

En relaciones lineales perfectas (es decir, cuando $r = \pm 1$), la recta que se ajusta exactamente a los datos tendrá de pendiente $\pm s_y/s_x$ (el signo de la pendiente es igual al signo del coeficiente de correlación) y pasa por el punto $(\bar{x}, \bar{y})$. Esta recta se denomina algunas veces **recta DT**. Nuestro ejemplo anterior sobre la temperatura muestra esta propiedad. Pero ¿qué sucede en situaciones en las que r no es exactamente ± 1 , es decir, cuando la relación entre X e Y no es perfectamente lineal?

Ejemplo 12.3.2 Arsénico en arroz

En la Sección 12.1 observamos un diagrama de dispersión que indicaba que la cantidad de arsénico en el arroz y de silicio en la paja del arroz parecían estar relacionadas linealmente ($r = -0,556$). La Figura 12.3.2 muestra un diagrama de dispersión de estos datos junto con la recta DT (línea discontinua). A primera vista, la recta DT parece ser un buen ajuste para estos datos. Sin embargo, una investigación más profunda sugiere otra cosa. Supongamos que deseáramos estimar la concentración media de arsénico en el arroz en plantas con concentraciones de silicio en la paja de 15 g/kg. La recta DT sugiere una concentración media de arsénico en el arroz de aproximadamente 190 $\mu\text{g/kg}$. Otra forma de estimar este valor sería simplemente utilizar la concentración media de arsénico en el arroz en las plantas de nuestra muestra que tienen concentraciones de silicio en la paja de aproximadamente 15 g/kg. La concentración media de arsénico para concentraciones de silicio en la paja entre 10 y 20 g/kg es de 158,6 $\mu\text{g/kg}$

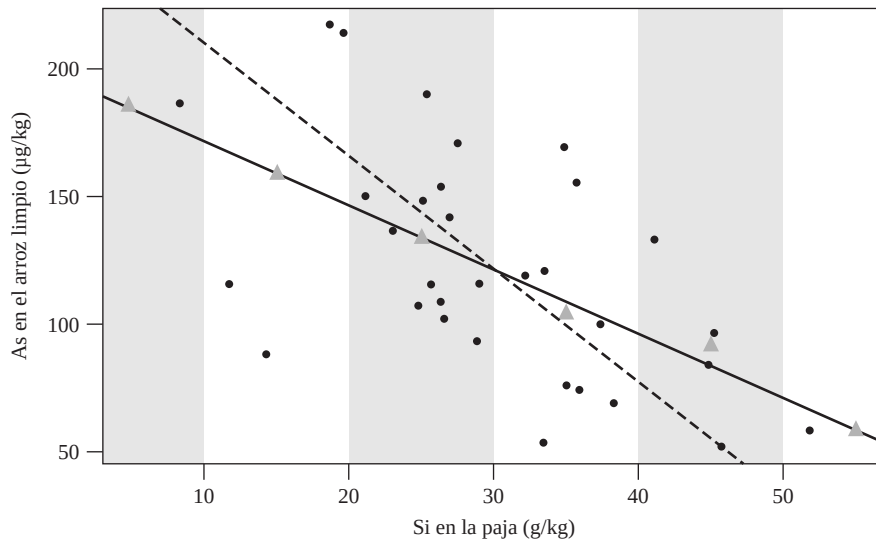


Figura 12.3.2 Concentraciones de arsénico en arroz frente a silicio en la paja para 32 plantas de arroz. Las líneas discontinua y continua son las rectas DT y de regresión ajustada, respectivamente. Los símbolos $\blacktriangle$ indican la concentración media de arsénico en el arroz en un intervalo de concentraciones de silicio en la paja especificado por el sombreado

(que se indica con el símbolo $\blacktriangle$ en la gráfica), que es considerablemente inferior al valor de 190 $\mu\text{g/kg}$ obtenido mediante la recta DT. De forma similar, para plantas con concentraciones de silicio en la paja de aproximadamente 45 g/kg, la recta DT indica un nivel de arsénico de aproximadamente 55 $\mu\text{g/kg}$, mientras que el nivel medio de arsénico en plantas con silicio entre 40 y 50 g/kg en nuestra muestra es de 91,4 $\mu\text{g/kg}$, un valor mucho mayor. ■

El ejemplo del arsénico en el arroz muestra que la recta DT tiende a sobreestimar el valor medio de Y para valores de X por debajo del promedio y a subestimar el valor medio de Y para valores de X por encima del promedio. La Figura 12.3.3 muestra un ejemplo incluso más exagerado para un conjunto de datos cuya correlación está más lejos de ± 1 y más cerca de cero ($r = -0,05$). Recuérdese que una correlación de cero indica que no existe relación lineal entre X e Y . Esta falta de correlación lineal se demuestra por el hecho de que el valor medio de Y es aproximadamente el mismo (≈ 17) independientemente del valor de X (la mayoría de los símbolos $\blacktriangle$ de la gráfica están cerca de 17).

Si la recta DT puede ser un resumen tan pobre, ¿por qué molestarse en estudiarla? Porque es un punto de inicio ideal que se basa en una relación lineal perfecta. Con una relación lineal perfecta (positiva) la recta DT es la recta de mejor ajuste y tiene una pendiente de s_y/s_x . Nuestros ejemplos ilustran que si la relación no es perfecta, entonces la relación entre los valores medios de Y y de X tiene una pendiente más plana. Matemáticamente, se puede demostrar que la recta que se ajusta mejor para predecir Y (en un cierto sentido), que se denomina recta de **regresión ajustada** o de **mínimos cuadrados**, tiene una pendiente igual a $r(s_y/s_x)$ y pasa por el punto $(\bar{x}, \bar{y})$. Es decir, para valores de Y que se alejen hacia la derecha una desviación típica del promedio, el valor de Y medio solo se alejará r desviaciones típicas del promedio hacia arriba (suponiendo que r es positivo; si r es negativo, entonces para valores de X que se alejen hacia la derecha una desviación típica del promedio, el valor medio de Y estará r desviaciones típicas por debajo del promedio).

Ejemplo 12.3.3 Arsénico en arroz

La Tabla 12.3.2 y la Figura 12.3.4 muestran un resumen y un diagrama de dispersión de los datos de arsénico en arroz. En este ejemplo, estimamos que las plantas con concentraciones de silicio en la paja que están $s_x = 10,04$ g/kg por encima del promedio (es decir, una desviación típica por encima del promedio) tendrán concentraciones de

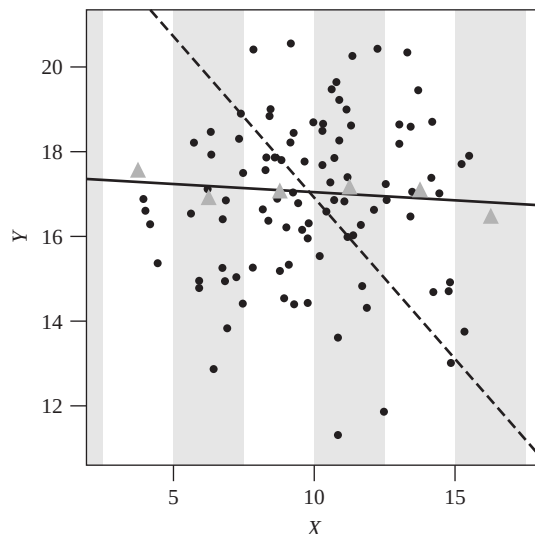


Figura 12.3.3 Diagrama de dispersión, recta DT (línea discontinua) y recta de regresión ajustada (línea continua) para una muestra de 100 valores de datos (x, y) con un coeficiente de correlación cercano a cero. Los símbolos $\blacktriangle$ indican los valores medios de Y en intervalos de valores de X especificados por el sombreado

arsénico en el arroz que están 25,19 $\mu\text{g/kg}$ por debajo del promedio ($r \times s_y = -0,566 \times 44,50 = -25,19$). De forma equivalente, la pendiente de la recta de regresión ajustada es

$$r(s_y/s_x) = -0,566 \times (44,50/10,04) = -2,51 \text{ (}\mu\text{g As/kg elevación)/(g Si/kg desplazamiento lateral)}$$

lo que significa que cada incremento adicional de 1 g/kg en la concentración de silicio en la paja está asociado con una disminución de 2,51 $\mu\text{g/kg}$ en la concentración de arsénico en el arroz, en promedio. ■

Tabla 12.3.2 Resumen de los datos de arsénico en arroz		
	$X = \text{Si en la paja (g/kg)}$	$Y = \text{As en el arroz (}\mu\text{g/kg)}$
Media	29,85	122,25
DT	10,04	44,50
	$r = -0,566$	

Ecuación de la recta de regresión

La ecuación de la recta se puede expresar como

$$Y = b_0 + b_1X$$

donde b_0 es la ordenada en el origen y b_1 es la pendiente de la recta. La pendiente b_1 es la tasa de cambio de Y con respecto a X .

La recta de regresión ajustada de Y con respecto a X se expresa como $\hat{y} = b_0 + b_1x$. Escribimos $\hat{y}$ (que se lee «Y-gorro») en vez de Y para recordar que esta recta solo proporciona valores de Y estimados o predichos. A menos que la correlación sea ± 1 , no podemos esperar que los valores de los datos estén exactamente sobre la recta. La recta de regresión ajustada estima el valor medio de Y para cualquier valor dado de X . Posteriormente comentaremos con mayor detalle este concepto de recta de regresión como *recta de promedios*.

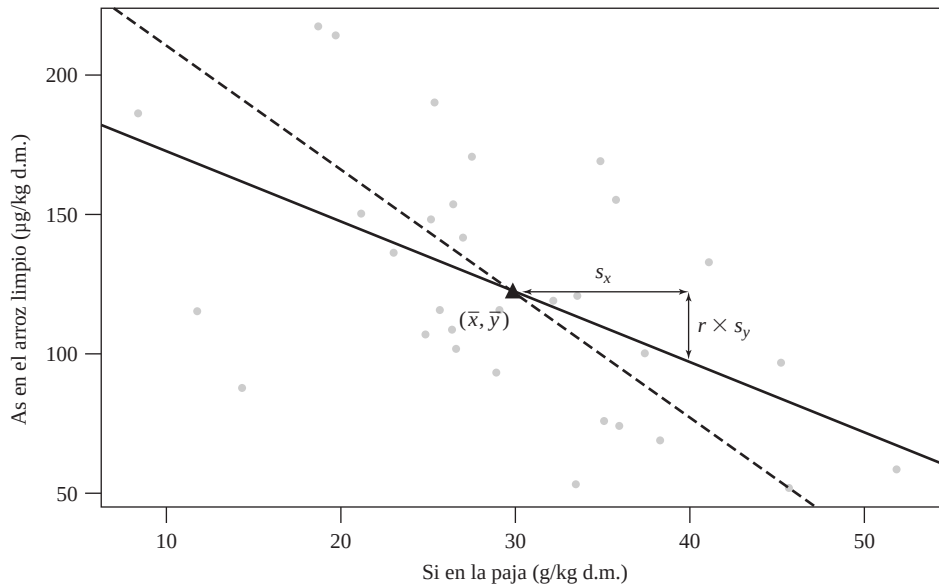


Figura 12.3.4 Concentraciones de arsénico en arroz frente a silicio en la paja para 32 plantas de arroz con la recta DT (línea discontinua) y la recta de regresión ajustada (en línea continua)

La pendiente y la ordenada en el origen de la recta de regresión de mínimos cuadrados* se calcula a partir de los datos como sigue:

Recta de regresión de mínimos cuadrados de Y con respecto a X

$$\text{Pendiente: } b_1 = r \left(\frac{s_y}{s_x} \right)$$

$$\text{Ordenada en el origen: } b_0 = \bar{y} - b_1 \bar{x}$$

Vimos previamente la motivación de la fórmula de la pendiente, b_1 . La fórmula de la ordenada en el origen es también fácil de motivar. Podemos volver a expresar la fórmula de la ordenada en el origen de Y como

$$\bar{y} = b_0 + b_1 \bar{x}$$

lo que demuestra que *la recta de regresión pasa por la media conjunta* $(\bar{x}, \bar{y})$ de nuestros datos.

Ilustraremos el uso de estas fórmulas continuando con nuestro ejemplo de arsénico en el arroz.

Ejemplo 12.3.4 Arsénico en el arroz

Obtuvimos previamente que la pendiente de la recta de regresión era $b_1 = r(s_y/s_x) = -2,15$ (μg de As/kg de arroz)/(g de Si/kg de paja). Utilizando este valor obtenemos la ordenada en el origen de Y ,

$$b_0 = 122,25 - (-2,51) \times 29,85 = 197,17 \mu\text{g/kg}$$

Por tanto, nuestra recta de regresión ajustada es $\hat{y} = 197,17 - 2,51x$, como se indicó previamente en la Figura 12.3.4. ■

* Hay otros métodos para obtener la recta de regresión ajustada. En este texto, consideramos solo la recta de regresión de mínimos cuadrados, que intenta minimizar las distancias verticales al cuadrado entre los valores de los datos y la recta ajustada.

Nótese que la ordenada en el origen de Y , el punto $(0, b_0) = (0, 197,17)$, no aparece en el diagrama de dispersión de la Figura 12.3.4 ya que los límites de la escala de la X no se extienden hasta cero, sino que varían entre 5 y 55 para producir una gráfica que se ajusta correctamente a los datos.

Gráfica de promedios

Si tenemos varias observaciones de Y a un nivel dado de X , podemos estimar la media poblacional de los valores de Y para el valor dado de X ($\mu_{Y|X}$) utilizando simplemente la media muestral de Y , $\bar{y}$, para ese valor dado de X . Podemos denominar esta media muestral como $\bar{y}|X^*$.

Algunas veces se puede calcular la media muestral, $\bar{y}$, para diversos conjuntos de valores de X . La gráfica de $\bar{y}|X$ se conoce como **gráfica de promedios**, ya que muestra los promedios (observados) de Y para diferentes valores de X .

Ejemplo 12.3.5 Anfetamina y consumo de alimentos

La Figura 12.3.5 es una gráfica de promedios para los datos de consumo de alimentos de la Tabla 12.1.1, que muestra los valores promedio de Y para tres niveles de X . Nótese que los tres $\bar{y}$ están casi en una recta. Esto da fundamento al uso de un lineal con estos datos. ■

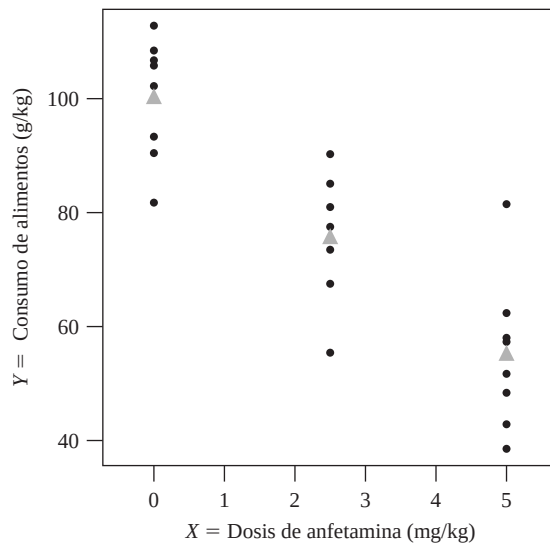


Figura 2.3.5 Gráfica de promedios (▲) para los datos de consumo de alimentos del Ejemplo 12.1.1 con los datos originales como puntos en color negro

Si los valores de $\bar{y}$ en la gráfica de promedios están exactamente sobre una recta, entonces esa recta es la recta de regresión y $\mu_{Y|X}$ se estima con $\bar{y}|X$. Sin embargo, en general, los valores de $\bar{y}$ no serán perfectamente colineales. En este caso, la recta de regresión es una versión *suavizada* de la gráfica de promedios, lo que produce un modelo ajustado en el que todas las estimaciones de $\mu_{Y|X}$ estarán en una recta. Al suavizar la gráfica de promedios en una recta, utilizamos información de *todas* las observaciones para estimar $\mu_{Y|X}$ para cualquier nivel de X .

Ejemplo 12.3.6 Anfetamina y consumo de alimentos

Si aplicamos las fórmulas de regresión anteriores a los datos sobre consumo de alimentos de la Tabla 12.1.1, obtenemos $b_0 = 99,3$ y $b_1 = -9,01$. Por tanto, la estimación de $\mu_{Y|X=0}$ es de 29,3 g/kg. Esta estimación difiere

* En la Sección 12.4 se puede encontrar una exposición más detallada de estas «medias condicionales».

ligeramente de $\bar{y}|X = 0$, que es 100,0 g/kg. El valor estimado de 99,3 hace uso de (1) los 8 valores de y cuando $X = 0$ (cuyo promedio es 100,0) y (2) la tendencia lineal establecida por los otros 16 puntos de datos, que mostraron un mayor consumo de alimento asociado a dosis menores. Asimismo, $\mu_{Y|X=25}$ es $99,3 - 9,01 \times 2,5 = 76,78$ g/kg, que difiere ligeramente de $\bar{y}|X = 25$ que es 75,5 g/kg y $\mu_{Y|X=5}$ es $99,3 - 9,01 \times 5 = 54,25$ g/kg, que difiere ligeramente de $\bar{y}|W = 5$ que es 55,0 g/kg. ■

La idea de suavizar la gráfica de promedios en una recta se traslada al escenario en el que tenemos solo una única observación para cada nivel de X , que es el caso del ejemplo del arsénico en el arroz. Cuando dibujamos una recta a través de un conjunto de datos (X, Y) , estamos expresando la confianza de que la dependencia subyacente del valor medio de Y con respecto a X sea suave, incluso aunque los datos muestren esa relación solo aproximadamente. La regresión lineal es una manera formal de proporcionar una descripción suave de los datos.

Suma de cuadrados residual

Consideremos ahora un estadístico que describe la dispersión de los puntos con respecto a la recta de regresión ajustada. La ecuación de la recta ajustada es $\hat{y} = b_0 + b_1x$. Por tanto, para cada valor x_i observando en nuestros datos hay un valor de Y predicho de

$$\hat{y}_i = b_0 + b_1x_i$$

Asociado también con cada pareja observada (x_i, y_i) hay un valor denominado **residuo**, definido como

$$e_i = y_i - \hat{y}_i$$

La Figura 12.3.6 muestra $\hat{y}$ y el residuo para un punto típico de datos (x_i, y_i) . Se puede demostrar que la suma de los residuos, teniendo en cuenta sus signos, es siempre cero debido al «equilibrado» de los puntos de datos por encima y por debajo de la recta de regresión ajustada. El *módulo* (valor absoluto) de cada residuo es la distancia vertical al punto del dato correspondiente desde el punto estimado por la recta ajustada.

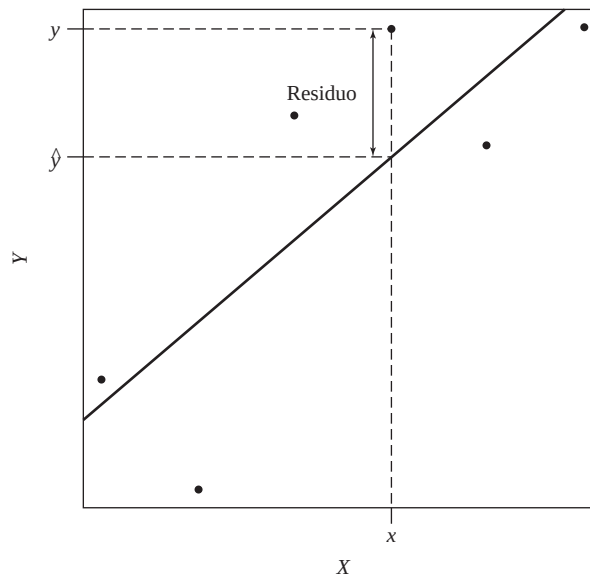


Figura 12.3.6 $\hat{y}$ y residuo de un punto de datos (x, y)

Nótese que el residuo se calcula en función de la distancia vertical. Al utilizar el modelo de regresión $\hat{y} = b_0 + b_1x$ estamos pensando que la variable X es un predictor y que la variable Y es la respuesta que depende de X . Nos preocupa principalmente lo cercano que cada valor observado, y_i , está de su valor predicho, $\hat{y}_i$. Por tanto, medimos distancias

verticales desde cada punto a la recta ajustada. Una medida resumen de las distancias de los puntos de datos a la recta de regresión es la **suma de cuadrados residual**, o **SC(resid)**, que se define como sigue:

Suma de cuadrados residual

$$SC(\text{resid}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$$

A partir de la definición resulta claro que la suma de cuadrados residual será pequeña si los puntos de datos están muy cerca de la recta.
 El ejemplo siguiente ilustra el cálculo de SC(resid).

Ejemplo 12.3.7 Arsénico en arroz

Para los datos de arsénico en arroz, la Tabla 12.3.3 indica cómo se calcularía SC(resid) a partir de su definición. Los valores se muestran abreviados para hacer la tabla más legible. ■

Tabla 12.3.3 Cálculo de SC(resid) para una parte de los datos de arsénico en arroz					
Obs #	x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
1	8,3	186,2	176,2...	10,0...	99,50...
2	11,8	115,5	167,6...	− 52,1...	2.716,00...
3	14,3	87,9	161,2...	− 73,3...	5.373,93...
4	18,7	217,2	150,2...	67,0...	4.492,74...
5	19,7	213,8	147,8...	66,0...	4.356,67...
6	21,2	150,0	144,0...	6,0...	35,53...
7	23,0	136,2	139,4...	− 3,2...	10,26...
8	25,1	148,3	134,1...	14,2...	200,46...
9	26,4	153,4	130,8...	22,6...	512,49...
⋮	⋮	⋮	⋮	⋮	⋮
27	38,3	69,0	101,0...	− 32,1...	1.028,99...
28	41,1	132,8	94,0...	38,8...	1.503,19...
29	45,2	96,6	83,6...	12,9...	167,11...
30	44,9	84,5	84,5...	0,0...	0,00...
31	45,7	51,7	82,5...	− 30,8...	948,51...
32	51,8	58,6	67,1...	− 8,5...	71,69...
Suma				0,0...	41.727,11 = SC(resid)

El criterio de mínimos cuadrados

Se han propuesto muchos criterios diferentes para definir la recta que «mejor» se ajusta a un conjunto de puntos de datos. El criterio clásico es el criterio de mínimos cuadrados:

Criterio de mínimos cuadrados

La «mejor» recta es la que minimiza la suma de cuadrados residual.

Las fórmulas dadas para b_0 y b_1 se obtuvieron mediante el criterio de mínimos cuadrados, aplicando cálculo para resolver el problema de minimización. (El desarrollo se presenta en el Apéndice 12.1). La recta de regresión ajustada se denomina también «recta de mínimos cuadrados».

El criterio de mínimos cuadrados puede parecer arbitrario e incluso innecesario. ¿Por qué no ajustar una recta a ojo con una regla? Realmente, a menos que los datos estén muy cerca de la recta, puede resultar sorprendentemente difícil ajustar una recta a ojo. El criterio de mínimos cuadrados proporciona una respuesta que no se basa en juicios individuales y que (como veremos en las Secciones 12.4 y 12.5) se puede interpretar realmente útil en términos de estimación de la distribución de los valores de Y para cada X fijo. Además, en la Sección 12.8 veremos que el criterio de mínimos cuadrados es un concepto versátil, cuyas aplicaciones van más allá que el simple ajuste de rectas.

La desviación típica residual

Un resumen de los resultados del análisis mediante regresión lineal debería incluir una medida de la cercanía de los puntos de datos a la recta ajustada. Una medida que se deriva de la suma de cuadrados residual y que es fácil de interpretar es la **desviación típica residual**, denominada s_e , que se define como sigue:

Desviación típica residual

$$s_e = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - 2}} = \sqrt{\frac{SC(\text{resid})}{n - 2}}$$

La desviación típica residual nos indica lo lejos que los puntos tienden a estar por encima o por debajo de la recta de regresión. Por tanto, dicha desviación especifica lo lejos que tienden a estar las predicciones realizadas mediante el modelo de regresión. Nótese que el factor del denominador es $n - 2$, en vez del habitual $n - 1$. El ejemplo siguiente ilustra el cálculo de s_e .

Ejemplo 12.3.8 Arsénico en arroz

Para los datos de arsénico en arroz, utilizamos $SC(\text{resid})$ del Ejemplo 12.3.7 para calcular

$$s_e = \sqrt{\frac{41.727,11}{32 - 2}} = \sqrt{1.390,90} = 37,30 \text{ } \mu\text{g/kg}$$

Por tanto, las predicciones de las concentraciones de arsénico en arroz basadas en el modelo de regresión tenderán a un error de aproximadamente 37,30 $\mu\text{g/kg}$ en promedio. ■

Nótese que la fórmula de s_e es muy parecida a la fórmula de s_y :

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}}$$

Ambas DT miden variabilidad en Y , pero la DT residual mide variabilidad con respecto a la *recta de regresión* y la DT ordinaria mide variabilidad con respecto a la media, $\bar{y}$. Hablando en términos generales, s_e es una medida de la distancia vertical de los puntos de los datos a la recta de regresión. (Nótese que la unidad de medida de s_e es la misma que la de Y , por ejemplo, $\mu\text{g/kg}$ en el caso de los datos de arsénico en arroz o gramos en el caso de los datos de serpientes del Ejemplo 12.2.1). La Figura 12.3.7 muestra el diagrama de dispersión y la recta de regresión para los datos de serpientes del Ejemplo 12.2.1 con los residuos representados como líneas verticales y la DT residual indicada como una recta vertical. Nótese que la DT residual indica aproximadamente la magnitud de un residuo típico. La obtención de la ecuación de esta recta y de la desviación típica residual aparece como un ejercicio al final de esta sección.

En muchos casos, s_e puede tener una interpretación cuantitativa más definida. Recordemos de la Sección 2.6 que en un conjunto de datos «adecuado», esperamos que aproximadamente el 68 % de las observaciones estén dentro de una DT alrededor de la media (y similarmente, para el 95 %, 2 DT). Recuérdesse también que esas reglas funcionan bien si los datos siguen aproximadamente una distribución normal. Se puede dar una interpretación similar para la DT residual: para

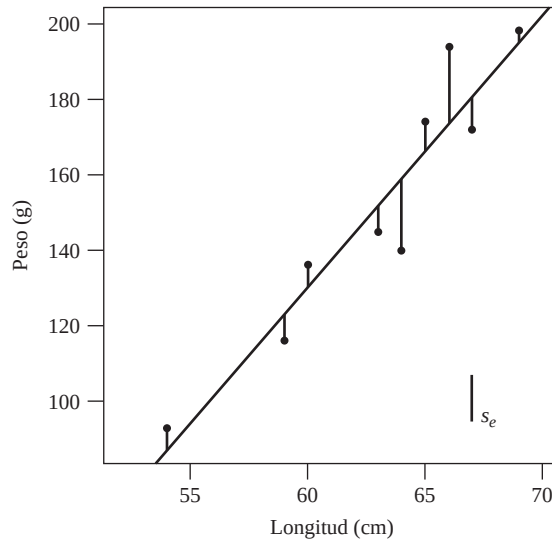


Figura 12.3.7 Peso frente a longitud de nueve serpientes mostrando los residuos y un segmento que muestra la magnitud de la DT residual

conjuntos de datos «adecuados» que no sean demasiado pequeños, podemos esperar que aproximadamente el 68 % de los valores de Y observados estén dentro de s_e alrededor de la recta de regresión. En otras palabras, esperamos que aproximadamente 68 % de los puntos de datos estén dentro de una distancia vertical de s_e por encima o por debajo de la recta de regresión (y similarmente para el 95 %, $\pm 2s_e$). Estas reglas funcionan mejor si los residuos siguen aproximadamente una distribución normal. Los datos de arsénico en arroz con los que hemos estado trabajando son adecuados para ilustrar la regla del 68 %.

Ejemplo 12.3.9 Arsénico en arroz

Para los datos de arsénico en arroz, la recta de regresión ajustada es $\hat{y} = 197,17 + 2,51x$ y la desviación típica residual es $s_e = 37,30$. La Figura 12.3.8 muestra los datos y la recta de regresión. Las rectas discontinuas están a una distancia vertical s_e alrededor de la recta de regresión. De los 32 puntos de los datos, 22 están entre las rectas discontinuas. Por tanto, $22/32$ o ≈ 69 % de los valores de Y observados están dentro de $\pm 1s_e$ de la recta de regresión. ■

El coeficiente de determinación

Hemos dicho que el módulo de r determina la fuerza de la relación lineal entre X e Y y hemos visto cómo se relaciona su valor con la pendiente de la recta de regresión. Cuando se eleva al cuadrado, proporciona también un resumen adicional e interpretable de la relación de regresión. El **coeficiente de determinación**, r^2 , describe la proporción de la varianza de Y explicada la relación lineal entre Y y X . Esta interpretación se sigue del siguiente hecho (demostrado en el Apéndice 12.2).

Propiedad 12.3.1: relación aproximada de r sobre s_e y s_y

El coeficiente de correlación r obedece a la siguiente relación aproximada

$$r^2 \approx \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

(La aproximación de la Propiedad 12.3.1 es mejor para valores de n grandes, pero se mantiene razonablemente bien para valores de n tan pequeños como 10). El numerador, $s_y^2 - s_e^2$, se puede interpretar aproximadamente como la varianza

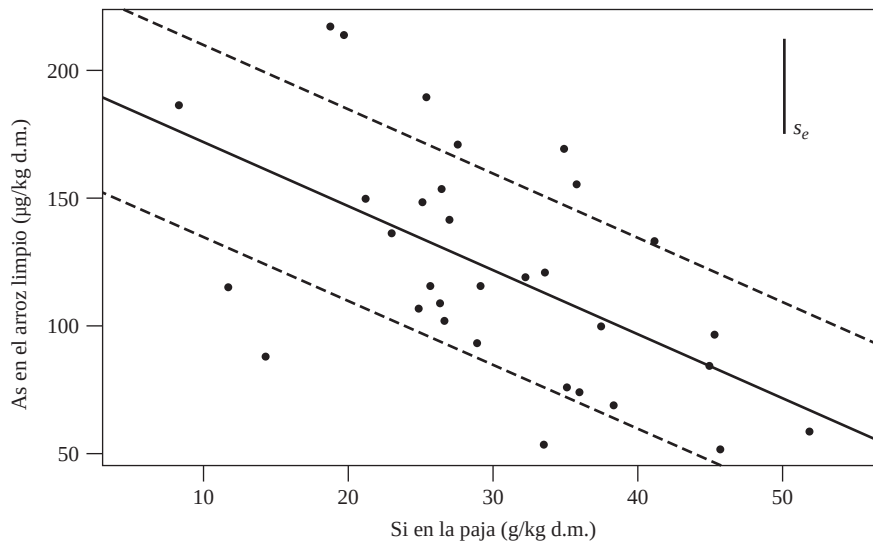


Figura 12.3.8 Arsénico en arroz frente a silicio en la paja para 32 plantas de arroz. Las rectas discontinuas están a una distancia vertical de s_e alrededor de la recta de regresión

total de Y explicada por la recta de regresión: es la diferencia entre la varianza de Y y la varianza residual, es decir, la varianza que queda después de ajustar la recta de regresión a los datos. Si la recta se ajusta muy bien a los datos, entonces s_e^2 será cercano a 0, por lo que este numerador será cercano a s_y^2 . En este caso r^2 estará cercano a 1. En el otro extremo, si la recta es un ajuste muy pobre, entonces s_e^2 estará muy cerca de s_y^2 y el numerador será cercano a 0. En este caso, por tanto, r^2 estará también cercano a 0. El denominador, s_y^2 , es la varianza de Y . Por tanto, la razón, r^2 , es la proporción de la varianza de Y explicada por la relación de regresión entre Y y X . Nótese que como $-1 \leq r \leq 1$, entonces $0 \leq r^2 \leq 1$. El ejemplo siguiente ilustra la interpretación y la aplicación de r^2 en el contexto.

Ejemplo 12.3.10 Arsénico en arroz

Para los datos de arsénico en arroz, obtuvimos $r = -0,566$, por lo que $r^2 = 0,320$ o el 32,0 %. Por tanto, el 32 % de la varianza de la concentración del arsénico en arroz es explicada por la relación lineal entre la concentración de arsénico en el arroz y la concentración de silicio en la paja. ■

Ejemplo 12.3.11 Anfetamina y consumo de alimentos

La desviación típica del consumo de alimentos para nuestra muestra completa de 24 ratas (es decir, combinando las ratas de las tres dosis de anfetamina) fue $s_y = 24,84$ g/kg. Supongamos además que r^2 se ha calculado y vale 0,739. ¿Cuál es la desviación típica estimada del consumo de alimentos para ratas que han recibido dosis de 4 mg/kg de anfetamina? Es decir, ¿cuál es el valor de $s_{Y|X=4}$?

Para responder a esta pregunta debemos darnos cuenta primero de que el valor de X es irrelevante. La desviación típica residual s_e describe la desviación típica de los valores de Y para cualquier valor dado de X y, por lo tanto, también para $X = 4$. Por consiguiente, necesitamos obtener el valor de s_e . Utilizando la Propiedad 12.3.1 tenemos

$$r^2 \approx 1 - \frac{s_e^2}{s_y^2}$$

Operando, obtenemos que la desviación típica (aproximada) del consumo de alimentos por ratas dado que han recibido una dosis de 4 mg/kg de anfetamina es

$$s_e \approx s_y \sqrt{1 - r^2} = 24,84 \sqrt{1 - 0,739} = 11,16 \text{ g/kg}$$

■

Ejercicios 12.3.1-12.3.10

12.3.1 En un estudio sobre síntesis de proteínas en el ovocito (célula del huevo en desarrollo) de la rana *Xenopus laevis*, un biólogo inyectó leucina marcada radiactivamente en ovocitos individuales. En varios momentos tras la inyección hizo medidas de radiactividad y calculó qué parte de la leucina se había incorporado a la proteína. Los resultados se muestran en la tabla que acompaña a este ejercicio. Cada valor de leucina es el contenido de leucina etiquetada en dos ovocitos. Todos los ovocitos provenían de la misma hembra¹³.

	Tiempo	Leucina
	0	0,02
	10	0,25
	20	0,54
	30	0,69
	40	1,07
	50	1,50
	60	1,74
Media	30,00	0,830
DT	21,60	0,637
$r = 0,993$ $SC(\text{resid}) = 0,035225$		

- Represente los datos. ¿Parece haber una relación entre X e Y ? ¿Es lineal o no lineal? ¿Débil o fuerte?
- Utilice regresión lineal para estimar la tasa de incorporación de la leucina etiquetada.
- Dibuje en su gráfica la recta de regresión.
- Calcule la desviación típica residual.

12.3.2 En una investigación sobre los efectos fisiológicos del alcohol (etanol) se asignaron aleatoriamente 15 ratones a tres grupos de tratamiento, y cada grupo recibió una dosis oral diferente de alcohol. Los niveles de dosificación fueron 1,5, 3,0 y 6,0 g de alcohol por kg de peso corporal. Se midió la temperatura corporal de cada ratón inmediatamente antes de que se le suministrara el alcohol y 20 minutos después de suministrárselo. La tabla que acompaña a este ejercicio muestra el descenso (antes menos después) en la temperatura corporal para cada ratón. (El valor negativo $-0,1$ se refiere a un ratón cuya temperatura se elevó en vez de descender)¹⁴.

Alcohol		Descenso en temperatura corporal (°C)					
Dosis (gm/kg)	Log(dosis) X	Valores individuales (Y)					
1,5	0,176	0,2	1,9	-0,1	0,5	0,8	0,66
3,0	0,477	4,0	3,2	2,3	2,9	3,8	3,24
6,0	0,778	3,3	5,1	5,3	6,7	5,9	5,26

- Represente el descenso medio de la temperatura corporal frente a la dosis. Represente el descenso medio de la temperatura corporal frente a $\log(\text{dosis})$. ¿Qué gráfica parece más lineal?
- Represente los puntos de datos (x, y) [siendo $X = \log(\text{dosis})$].
- Para la regresión de Y con respecto a $X = \log(\text{dosis})$, los cálculos preliminares produjeron los siguientes resultados: $\bar{x} = 0,477$, $\bar{y} = 3,05333$, $s_x = 0,25439$, $s_y = 2,13437$, $r = 0,91074$. Calcule la recta de regresión ajustada y la desviación típica residual (aproximada).
- Dibuje la recta de regresión sobre su gráfica.
- ¿Es este estudio un ejemplo de estudio observacional o un experimento? ¿Cómo puede decirlo?
- ¿Podrían utilizarse los datos de este estudio para determinar si el alcohol disminuye o no la temperatura corporal? Explique brevemente su respuesta.

12.3.3 Considere los datos sobre pesos de mazorcas del Ejercicio 12.2.5.

- Utilice los resúmenes del Ejercicio 12.2.5 para calcular la recta de regresión ajustada y la desviación típica del residuo aproximada.
- Interprete el valor de la pendiente de la recta de regresión, b_1 , en el contexto de este ejercicio.
- $SC(\text{resid}) = 1.337,3$. Utilice este valor para calcular la desviación típica del residuo. ¿Cómo se compara con el valor aproximado determinado en el apartado (a)?
- Interprete el valor de s_e en el contexto de este ejercicio.
- ¿Qué proporción de la variación del peso de la mazorca es explicado por la relación lineal entre el peso de la mazorca y la densidad?

12.3.4 Considere los datos sobre crecimiento de hongos del Ejercicio 12.2.6.

- Calcule la regresión lineal de Y con respecto a X .
- Dibuje los datos y añada a su gráfica la recta de regresión. ¿Parece que la recta se ajusta bien a los datos?
- $SC(\text{resid}) = 16,7812$. Utilice este dato para calcular s_e . ¿En qué unidades está s_e ?
- Dibuje una recta en su gráfica que represente la magnitud de s_e . (Véase la Figura 12.3.8).

12.3.5 Considere los datos sobre gasto de energía del Ejercicio 12.2.7.

- Calcule la regresión lineal de Y con respecto a X .
- Dibuje los datos y añada a su gráfica la recta de regresión. ¿Parece que la recta se ajusta bien a los datos?
- Interprete el valor de la pendiente de la recta de regresión, b_1 , en el contexto de este ejercicio.
- $SC(\text{resid}) = 21.026,1$. Utilice este dato para calcular s_e . ¿En qué unidades está s_e ?
- Dibuje una recta en su gráfica que represente la magnitud de s_e . (Véase la Figura 12.3.8).

12.3.6 El serbal (*Sorbus aucuparia*) es un árbol que crece en un amplio intervalo de altitudes. Para estudiar cómo se adapta el árbol a sus hábitats variables, unos investigadores recogieron ramitas con yemas en 12 árboles que crecían en varias altitudes en North Angus, Escocia. Las yemas se llevaron al laboratorio y se midió su tasa de respiración en oscuridad. La tabla que acompaña a este ejercicio muestra la altitud en origen (en metros) de cada grupo de yemas y la tasa de respiración en oscuridad (expresada como μl de oxígeno por hora por mg de peso de tejido en seco)¹⁵.

	Altitud de origen X (m)	Tasa de respiración Y ($\mu\text{l}/\text{h} \times \text{mg}$)
	90	0,11
	230	0,20
	240	0,13
	260	0,15
	330	0,18
	400	0,16
	410	0,23
	550	0,18
	590	0,23
	610	0,26
	700	0,32
	790	0,37
Media	433,333	0,21000
DT	214,617	0,07710
$r = 0,88665$ $SC(\text{resid}) = 0,013986$		

- Calcule la regresión lineal de Y con respecto a X .
- Dibuje los datos y la recta de regresión.
- Interprete el valor de la pendiente de la recta de regresión, b_1 , en el contexto de este ejercicio.
- Calcule la desviación típica residual.

12.3.7 Unos científicos estudiaron la relación entre la longitud del cuerpo de una rana toro y su distancia máxima de salto. El estudio incluyó once ranas toro. Los resultados se muestran en la tabla¹⁶.

- Calcule la regresión lineal de Y con respecto a X .
- Interprete el valor de la pendiente de la recta de regresión, b_1 , en el contexto de este ejercicio.
- ¿Qué proporción de la variación de las distancias máximas de salto se puede explicar por la relación lineal entre la distancia de salto y la longitud de la rana?
- Calcule la desviación típica residual y especifique sus unidades.
- Interprete el valor de la desviación típica residual en el contexto de este ejercicio.

Rana toro	Longitud X (mm)	Salto máximo Y (cm)
1	155	71,0
2	127	70,0
3	136	100,0
4	135	120,0
5	158	103,3
6	145	116,0
7	136	109,2
8	172	105,0
9	158	112,5
10	162	114,0
11	162	122,9
Media	149,6364	103,9909
DT	14,4725	17,9415
$r = 0,28166$ $SC(\text{resid}) = 2.963,61$		

12.3.8 El pico de velocidad de flujo de una persona es la máxima velocidad a la que dicha persona puede expeler aire después de tomar una respiración profunda. El pico de velocidad de flujo se mide en unidades de litros por minuto y da una indicación de la salud respiratoria de la persona. Unos investigadores midieron el pico de velocidad de flujo y la altura de una muestra de 17 hombres. Los resultados se presentan en la tabla¹⁷.

Sujeto	Altura X (cm)	Pico de velocidad de flujo Y (l/min)
1	174	733
2	183	572
3	176	500
4	169	738
5	183	616
6	186	787
7	178	866
8	175	670
9	172	550
10	179	660
11	171	575
12	184	577
13	200	783
14	195	625
15	176	470
16	176	642
17	190	856
Media	180,4118	660,0000
DT	8,5591	117,9952
$r = 0,32725$ $SC(\text{resid}) = 198,909$		

- (a) Calcule la regresión lineal de Y con respecto a X .
- (b) ¿Qué proporción de la variación de la velocidad de flujo se puede explicar por la regresión lineal de la velocidad de flujo con respecto a la altura?
- (c) Para cada sujeto, calcule el pico de velocidad de flujo predicho utilizando la ecuación de regresión del apartado (a).
- (d) Para cada sujeto, calcule el residuo utilizando los resultados del apartado (c).
- (e) Calcule s_e y especifique sus unidades.
- (f) ¿Qué porcentaje de los puntos de datos están dentro de $\pm s_e$ alrededor de la recta de regresión? Es decir, ¿qué porcentaje de los 17 residuos están en el intervalo $(-s_e, s_e)$?

12.3.9 Para cada uno de los siguientes conjuntos de datos, prepare una gráfica como la de la Figura 12.3.8, que muestre

los datos, la recta de regresión ajustada y dos rectas cuya distancia vertical por encima y por debajo de la recta de regresión sea s_e . ¿Qué porcentaje de los puntos de datos están dentro de $\pm s_e$ alrededor de la recta de regresión? ¿Qué porcentaje de los puntos de datos espera encontrar dentro de $\pm s_e$ alrededor de la recta de regresión? ¿Cómo se comparan estos valores?

- (a) Los datos de temperatura corporal del Ejercicio 12.3.2.
- (b) Los datos sobre cosechas de maíz del Ejercicio 12.3.3.

12.3.10 Suponga que se utiliza una muestra grande de parejas (x, y) para ajustar una regresión de Y con respecto a X . Suponga ahora que observamos 100 parejas (x, y) más. ¿Aproximadamente cuántas de estas nuevas observaciones se podría esperar que se alejarán más de $2s_e$ de la recta de regresión?

12.4 Interpretación paramétrica de la regresión: el modelo lineal

Uno de los usos de análisis de regresión es simplemente proporcionar una descripción concisa de los datos. Los valores b_0 y b_1 localizan la recta de regresión y s_e describe la dispersión de los puntos alrededor de dicha recta.

Sin embargo, en muchos casos, esa descripción de los datos no es suficiente. En esta sección consideraremos la realización de inferencias en una población más grande a partir de los datos. En capítulos anteriores hemos hablado de una o varias poblaciones de valores de Y . Ahora, para incorporar la variable X también, necesitamos ampliar la noción de población.

Poblaciones condicionales y distribuciones condicionales

Una **población condicional** de valores de Y es una población de valores de Y asociados con un valor fijo, o dado, de X . Dentro de una población condicional podemos hablar de la **distribución condicional** de Y . La media y la desviación típica de una distribución poblacional condicional se denominan

$\mu_{Y|X}$ = Valor de la media poblacional de Y para un X dado

$\sigma_{Y|X}$ = DT poblacional de los valores de Y para un X dado

(Nótese que el símbolo « $|$ » de «dado» es el mismo utilizado para la probabilidad condicional en los Capítulos 3 y 10). El ejemplo siguiente ilustra esta notación.

Ejemplo 12.4.1 Anfetamina y consumo de alimentos

En el experimento con ratas presentado en el Ejemplo 12.1.1, la variable de respuesta Y es el consumo de alimentos y los tres valores de X (dosis) eran $X = 0$, $X = 2,5$ y $X = 5$. En el Ejemplo 12.3.5 examinamos la gráfica de promedios y consideramos los datos de consumo de alimentos como tres muestras independientes (como en un ANOVA). En el contexto del ANOVA denotamos las tres medias poblacionales como μ_1 , μ_2 y μ_3 . En notación de análisis de regresión, estas medias se denotarían como

$$\mu_{Y|X=0} \quad \mu_{Y|X=2,5} \quad \mu_{Y|X=5}$$

respectivamente. De forma similar, las tres desviaciones típicas poblacionales, que se denotaron como σ_1 , σ_2 y σ_3 en un contexto de ANOVA, se denotarían como

$$\sigma_{Y|X=0} \quad \sigma_{Y|X=2,5} \quad \sigma_{Y|X=5}$$

respectivamente. En otras palabras, los símbolos

$$\mu_{Y|X} \quad Y \quad \sigma_{Y|X}$$

representan la media y desviación típica de los valores del consumo de alimentos para ratas a las que se les ha dado la dosis X de anfetamina. ■

En estudios observacionales, las distribuciones condicionales pertenecen a subpoblaciones en vez de a grupos de tratamiento experimentales, como en el ejemplo siguiente.

Ejemplo 12.4.2 Alturas y pesos de hombres jóvenes

Consideremos las variables

$$X = \text{Altura}$$

e

$$Y = \text{Peso}$$

de una población de hombres jóvenes. Las medias y desviaciones típicas condicionales son

$$\mu_{Y|X} = \text{Peso medio de hombres cuya altura es de } X \text{ pulgadas}$$

$$\sigma_{Y|X} = \text{DT de los pesos de los hombres cuya altura es de } X \text{ pulgadas}$$

Por tanto, $\mu_{Y|X}$ y $\sigma_{Y|X}$ son la media y la desviación típica de los pesos en la subpoblación de hombres cuya altura es X . Por supuesto, hay una subpoblación diferente para cada valor de X . ■

El modelo lineal

Cuando realizamos un análisis de regresión lineal, pensamos que Y tiene una distribución que depende de X . El análisis puede tener una interpretación paramétrica si se cumplen dos condiciones. Estas condiciones, que constituyen el **modelo lineal**, se presentan en el recuadro siguiente.

El modelo lineal

1. *Linealidad.* $Y = \mu_{Y|X} + \varepsilon$, siendo $\mu_{Y|X}$, es una función lineal de X . Es decir

$$\mu_{Y|X} = \beta_0 + \beta_1 X$$

Por tanto, $Y = \beta_0 + \beta_1 X + \varepsilon$.

2. *Desviación típica constante.* $\sigma_{Y|X}$ no depende de X . Denominaremos σ_ε a este valor constante.

En el modelo lineal $Y = \beta_0 + \beta_1 X + \varepsilon$, el término ε representa un **error aleatorio**. Incluimos este término en el modelo para reflejar el hecho de que Y varía, incluso cuando X es fija. La variabilidad de Y para un valor fijo de X se mide mediante la desviación típica condicional de Y , $\sigma_{Y|X}$. Pero, debido a que el modelo lineal estipula que esta desviación típica es la misma para cada valor de X , utilizaremos comúnmente la notación σ_ε para representar esta desviación típica y nos referiremos a ella como la desviación típica del error aleatorio.

Los dos ejemplos siguientes muestran el significado del modelo lineal.

Ejemplo 12.4.3 Anfetamina y consumo de alimentos

Para el experimento de consumo de alimentos con ratas, el modelo lineal asevera que (1) la media poblacional del consumo de alimentos es una función lineal de la dosis y qué (2) los valores de la desviación típica poblacional del consumo de alimentos son los mismos para todas las dosis. Nótese que la segunda condición es análoga a la condición del ANOVA de que las DT poblacionales son iguales: $\sigma_1 = \sigma_2 = \sigma_3$. El modelo lineal permite también que haya variabilidad en Y cuando X es fija. Por ejemplo, había 8 observaciones para las que $X = 5$. Los 8 valores de y tienen un promedio de 55,0, pero ninguna de las observaciones era igual a 55,0. Había sustancial variabilidad dentro de los 8 valores de y . La variabilidad queda cuantificada por la DT de 13,3. ■

Ejemplo 12.4.4 Alturas y pesos de hombres jóvenes

Consideremos una población ficticia idealizada de hombres jóvenes cuya distribución conjunta de pesos y alturas se ajusta exactamente al modelo lineal. En nuestra población ficticia asumiremos que las medias y DT condicionales de los pesos dada una altura fija son las siguientes:

$$\begin{aligned}\mu_{Y|X} &= -145 + 4,25X \\ \sigma_{\varepsilon} &= 20\end{aligned}$$

Por tanto, los parámetros de la regresión en la población son $\beta_0 = -145$ y $\beta_1 = 4,25$ (esta población ficticia se parece a los jóvenes estadounidenses de 17 años)¹⁸. Por tanto, el modelo es $Y = -145 + 4,25X + \varepsilon$.

La Tabla 12.4.1 muestra las medias condicionales y DT de $Y =$ peso para unos cuantos valores seleccionados de $X =$ altura. La Figura 12.4.1 muestra las distribuciones condicionales de Y dado X para esas subpoblaciones seleccionadas.

Tabla 12.4.1 Medias y DT condicionales de pesos y alturas dados en una población de hombres jóvenes*		
Altura (plg) X	Peso medio (lb) $\mu_{Y X}$	Desviación típica de los pesos (lb) $\sigma_{Y X}$
64	127	20
68	144	20
72	161	20
76	178	20

* Nótese que todos los valores de $\sigma_{Y|X}$ son los mismos; son iguales a $\sigma_{\varepsilon} = 20$

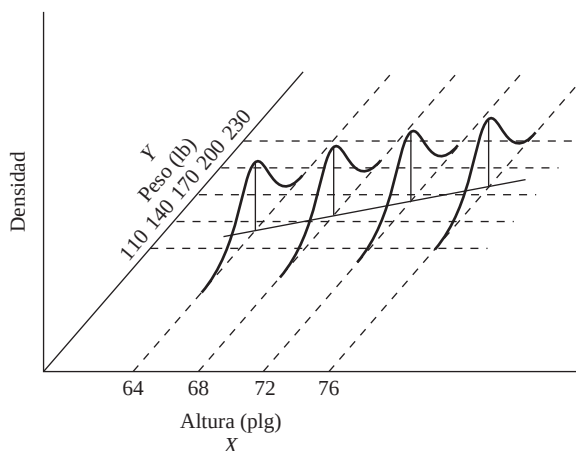


Figura 12.4.1 Distribuciones condicionales de pesos y alturas dados en una población de hombres jóvenes

Nótese, por ejemplo que si altura = 68 (plg), entonces el peso medio es 144 (lb) y la DT de los pesos es 20 (lb). Para esta subpoblación, $Y = 144 + \varepsilon$. Si un hombre joven concreto cuya altura es de 68 pulgadas pesa 145 libras, entonces para él $\varepsilon = 1$. Si otro hombre joven con 68 pulgadas de altura pesa 140 libras, entonces en su caso $\varepsilon = -4$. Por supuesto, β_0 , β_1 y ε no son en general observables. Este ejemplo es ficticio. ■

Observación. Realmente, el término *regresión* no se limita a la regresión lineal. En general, una relación entre $\mu_{Y|X}$ y X se denomina *regresión de Y con respecto a X*. El supuesto de linealidad afirma que la regresión de Y con respecto a X es lineal en vez de, por ejemplo, una función curvilínea.

Estimación en el modelo lineal

Consideremos ahora el análisis de un conjunto de datos (X, Y) . Supongamos que asumimos que el modelo lineal es una descripción adecuada de la verdadera relación de Y y X . Supongamos además que deseamos adoptar el siguiente **modelo de submuestreo aleatorio**:

Modelo de submuestreo aleatorio

Para cada pareja observada (x, y) , consideramos que el valor de y ha sido muestreado aleatoriamente de la población condicional de valores de Y asociados con el valor x de X .

Dentro del marco del modelo lineal y del modelo del submuestreo aleatorio, los valores b_0 , b_1 , y s_e calculados mediante un análisis de regresión se pueden interpretar como estimadores de los parámetros poblacionales:

b_0 es un estimador de β_0

b_1 es un estimador de β_1

s_e es un estimador de σ_e

Ejemplo 12.4.5 Longitud y peso de serpientes

A partir de los resúmenes de los datos de serpientes del Ejemplo 12.2.1 y 12.2.2, se pueden calcular los siguientes coeficientes de regresión: $b_0 = -301$, $b_1 = 7,1$ y $s_e = 12,5$ (el cálculo de dichos coeficientes a partir de los resúmenes que se proporcionan sería un buen ejercicio). Por tanto,

-301 es nuestro estimador de β_0

$7,19$ es nuestro estimador de β_1

$12,5$ es nuestro estimador de σ_e ■

La aplicación del modelo lineal a los datos de las serpientes ha producido dos beneficios. Primero, la pendiente de la recta de regresión, $7,19$ g/cm, es una estimación del parámetro morfológico («peso por unidad de longitud»), que es de potencial interés biológico para caracterizar la población de serpientes. Segundo, hemos obtenido una estimación ($12,5$ g) de la variabilidad de peso entre serpientes de longitud fija, incluso aunque no se pueda estimar directamente esta variabilidad porque ninguna pareja de serpientes observadas tuvieran la misma longitud.

Interpolación en el modelo lineal

En la Sección 12.3 consideramos la recta de regresión como una recta de promedios. La idea de suavizar la gráfica de los promedios con una línea recta se puede ampliar a un escenario en el que tenemos una única observación en cada nivel de X . Al dibujar una recta a través de un conjunto de datos (X, Y) , estamos expresando la creencia de que la dependencia subyacente de Y con respecto a X es suave, incluso aunque los datos solo muestren esa relación aproximadamente. La regresión lineal es una manera formal de proporcionar una descripción suave de los datos, como se ilustra en el siguiente ejemplo.

Ejemplo 12.4.6 Arsénico en arroz

¿Cuál es la media y la desviación típica de concentraciones de arsénico en arroz para plantas con concentraciones de silicio en la paja de 33 g/kg? Ninguna de nuestras plantas observadas tiene una concentración de silicio en la paja de 33 g/kg. Si hubiera algunas observaciones con ese nivel de silicio, podríamos promediar las concentraciones de arsénico asociadas para obtener la respuesta a esta pregunta, pero como hay una relación lineal aparente entre X e Y , podemos utilizar la recta para obtener una estimación incluso mejor de la concentración media de arsénico en arroz que usa todos los datos. En el Ejemplo 12.3.4 obtuvimos que la ecuación de regresión era $\hat{y} = 197,17 - 2,51x$ y $s_e = 37,30$. Por tanto, la concentración de arsénico media estimada para paja con 33 g/kg de silicio es $197,17 - 2,51 \times 33 = 114,35$ $\mu\text{g/kg}$, con una desviación típica de $s_e = 37,30$ $\mu\text{g/kg}$. La Figura 12.4.2 muestra la interpolación gráficamente. ■

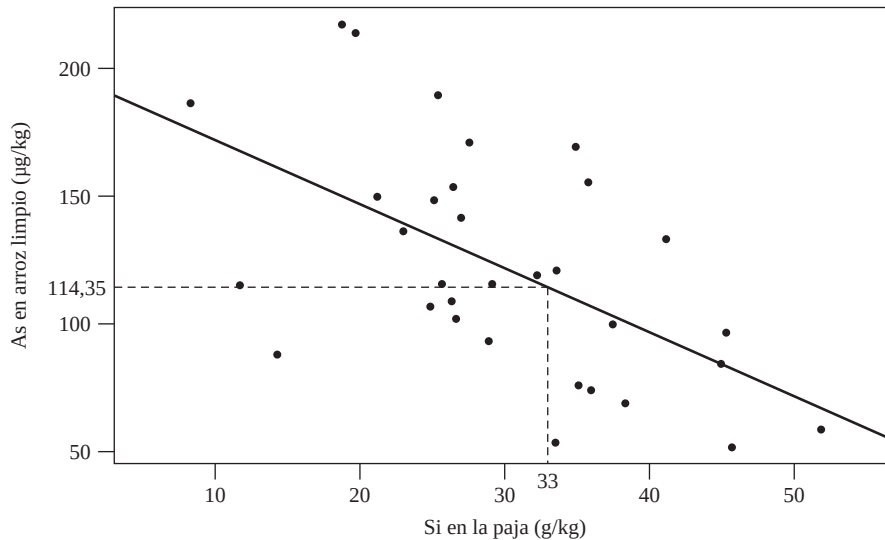


Figura 12.4.2 Concentraciones de arsénico en arroz frente a silicio en la paja para 32 plantas de arroz

Nótese que la estimación de la media utiliza la condición de linealidad del modelo lineal, mientras que la estimación de la desviación típica utiliza la condición de desviación típica constante. En algunas situaciones solo podría ser creíble la condición de linealidad, y entonces solo se podría estimar la media.

El Ejemplo 12.4.6 es un ejemplo de **interpolación**, porque los valores de X escogidos ($X = 33$ para el arsénico en el arroz) está dentro del intervalo de valores observados de X . Por el contrario, la **extrapolación** es el uso de una recta de regresión (o de otra curva) para predecir valores de Y para valores de X que están fuera del intervalo de los datos. La extrapolación debería evitarse siempre que sea posible, porque en general no hay seguridad de que la relación entre $\mu_{Y|X}$ y X siga siendo lineal para valores de X fuera del intervalo observado. Muchas relaciones biológicas son lineales solo en una parte de los posibles intervalos de valores de X . A continuación sigue un ejemplo.

Ejemplo 12.4.7 Anfetamina y consumo de alimentos

La relación dosis-respuesta para el experimento de consumo de alimentos por ratas del Ejemplo 12.1.1 tiene el aspecto aproximado que muestra la Figura 12.4.3¹⁹. Los datos cubren solo la parte lineal de la relación. Es claro que sería erróneo extrapolar la recta ajustada a $X = 10$ o $X = 15$. ■

Predicción y modelo lineal

Consideremos el escenario de utilizar la altura, X , para predecir el peso, Y , en un grupo grande de hombres jóvenes para los que el peso medio es 150 libras. Supongamos que se escoge aleatoriamente un hombre joven y debemos predecir su peso.

1. Si no sabemos nada sobre la altura del hombre, entonces el mejor estimador que podemos dar de su peso es el peso medio global, $\bar{y} = 150$.
2. Supongamos que sabemos que la altura del hombre es de 76 pulgadas. Si conocemos que el peso medio de todos los hombres del grupo cuya altura de 76 pulgadas es de 180 libras, entonces podemos utilizar el promedio condicional, $\bar{y}|x = 76$, como predicción del peso del hombre. Esperamos que esta predicción, que esencialmente es utilizar la gráfica de promedios (pero sin suavizar) sea más exacta que la dada en el apartado 1.
3. Supongamos que sabemos que la altura del hombre es de 76 pulgadas y que también sabemos que la ecuación de regresión de mínimos cuadrados es $Y = -140 + 4,3X$. Entonces podemos utilizar el valor $x = 76$ para obtener una predicción, que sería $-140 + 4,3 \times 76 = 186,8$.

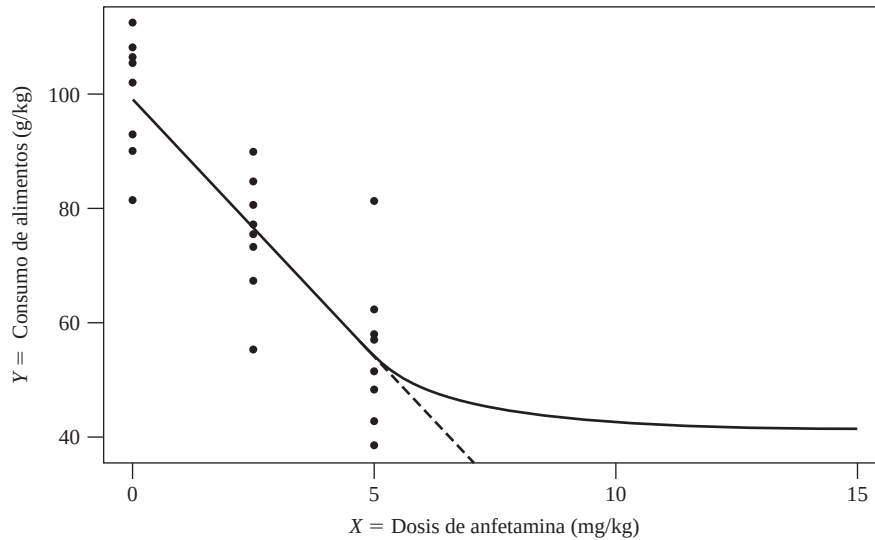


Figura 12.4.3 Curva de datos-respuesta (en respuesta media frente a dosis) del experimento de consumo de alimentos por ratas

¿Es la predicción del apartado 3 mejor que la del apartado 2? Como usar la ecuación de regresión implica suavizar la gráfica de promedios, debemos esperar que la predicción del apartado 3 sea mejor que la predicción del apartado 2 *en tanto en cuanto creemos que existe una relación lineal entre altura y peso*. La predicción del apartado 3 tiene la ventaja de utilizar información de todos los puntos de datos, no solo aquellos para los que $x = 76$. El método del apartado 3 tiene también la ventaja de que se pueden realizar predicciones cuando el valor de x (la altura) no es ninguno de los del conjunto de datos original (como se presentó en la subsección anterior «Interpolación en el modelo lineal»), por lo que no es conocido. Sin embargo, el método del apartado 3 producirá predicciones pobres si la relación lineal no es cierta. Por tanto, es muy importante pensar sobre las relaciones, y explorarlas gráficamente antes de utilizar un modelo de regresión.

Ejercicios 12.4.1-12.4.9

12.4.1 En los datos del Ejercicio 12.2.6 había dos observaciones para las que $X = 0$. La respuesta media (valor de Y) de esos puntos es $\frac{33,3 + 31,0}{2} = 32,15$. Sin embargo, la ordenada en el origen de la recta de regresión, b_0 , no es 32,15. ¿Por qué no? ¿Por qué es b_0 , en vez de 32,15, un estimador mejor del crecimiento medio de hongos cuando la concentración de ácido laetisárico es cero?

12.4.2 En referencia a los datos de temperatura corporal del Ejercicio 12.3.2, suponiendo que el modelo lineal es aplicable, estime la media y la desviación típica del descenso en temperatura corporal que se observaría en un ratón dada una dosis de alcohol de 2 g/kg. [Sugerencia: ¿es la variable X dosis o $\log(\text{dosis})$?].

12.4.3 En referencia a los datos de pesos de mazorcas de los Ejercicios 12.2.5 y 12.3.3, suponga que el modelo lineal es aplicable.

- Estime el peso medio de la mazorca que se esperaría en una parcela que contuviera (i) 100 plantas; (ii) 120 plantas.
- Suponga que cada planta produce una mazorca. ¿Cuánto grano se esperaría obtener de una parcela que contuviera (i) 100 plantas; (ii) 120 plantas?

12.4.4 (Continuación del Ejercicio 12.4.3). Para los datos de pesos de mazorcas, $SC(\text{resid}) = 1.337,3$. Estime la desviación típica del peso de las mazorcas en parcelas que contienen (i) 100 plantas; (ii) 120 plantas.

12.4.5 En referencia a los datos sobre crecimiento de hongos del Ejercicio 12.2.6, para dichos datos, $SC(\text{resid}) = 16,7812$. Suponiendo que el modelo lineal es aplicable, obtenga estimaciones de la media y la desviación típica del crecimiento de hongos con una concentración de ácido laetisárico de 15 $\mu\text{g/ml}$.

12.4.6 En referencia a los datos sobre gasto de energía del Ejercicio 12.2.7, suponiendo que el modelo lineal es aplica-

ble, estime el gasto de energía durante 24 horas de un hombre cuya masa libre de grasa es de 55 kg.

12.4.7 En referencia a la actividad de la bomba de Ca del Ejercicio 12.2.8, para esos datos $SC(\text{resid}) = 21.984,623$.

- (a) Suponiendo que el modelo lineal es aplicable, estime la media y la desviación típica de la actividad de la bomba de Ca basal para niños nacidos de madres con un nivel de Hg en el pelo de 3 $\mu\text{g/g}$.
- (b) Utilizando los valores calculados en el apartado (a) para fundamentar su respuesta, ¿sería sorprendente que una madre con un nivel de Hg en el pelo de 3 $\mu\text{g/g}$ diera a

luz un niño con actividad de la bomba de Ca basal por encima de 4.000 nmol/mg/h?

12.4.8 En referencia a los datos sobre las ranas toro del Ejercicio 12.3.7, suponiendo que el modelo lineal es aplicable, estime la máxima longitud de salto de una rana toro cuya longitud corporal sea de 150 mm.

12.4.9 En referencia a los datos de flujo de pico del Ejercicio 12.3.8, suponiendo que el modelo lineal es aplicable, obtenga estimaciones de la media y la desviación típica del flujo de pico para hombres cuya altura sea de 180 cm.

12.5 Inferencia estadística con respecto a β_1

El modelo lineal proporciona interpretaciones de b_0 , b_1 y ε que permiten ir más allá de la descripción de datos y entrar en el dominio de la inferencia estadística. En esta sección consideramos la inferencia sobre la verdadera pendiente β_1 de la recta de regresión. Los métodos se basan en la condición de que la distribución poblacional condicional de Y para cada valor de X es una distribución normal. Esto es equivalente a plantear que en el modelo lineal de $Y = \beta_0 + \beta_1 X + \varepsilon$, los valores de ε provienen de una distribución normal.

El error típico de b_1

Dentro del contexto de modelo lineal, b_1 es un estimador de β_1 . Como todos los estimadores calculados a partir de los datos, b_1 está sujeto al error de muestreo. El error típico de b_1 se calcula como sigue:

Error típico de b_1

$$ET_{b_1} = \frac{s_e}{s_x \sqrt{n-1}}$$

El ejemplo siguiente ilustra el cálculo de ET_{b_1} .

Ejemplo 12.5.1 Longitud y peso de serpientes

Para los datos de las serpientes, obtuvimos en la Tabla 12.2.2 que $n = 9$, $s_x = 4,637$, y en el Ejemplo 12.4.5 que $s_e = 12,5$. El error típico de b_1 es

$$ET_{b_1} = \frac{12,5}{4,637 \sqrt{9-1}} = 0,9531$$

Para resumir, la pendiente de la recta de regresión ajustada (del Ejemplo 12.4.5) es

$$b_1 = 7,19 \text{ g/cm}$$

y el error típico de esta pendiente es

$$ET_{b_1} = 0,95 \text{ g/cm}$$

Estructura del ET. Veamos cómo depende el error típico de b_1 de varios aspectos de los datos. De la misma forma que $ET_{\bar{y}}$ depende de la variabilidad de los datos de $Y(s_y)$ y del tamaño de la muestra (n), ET_{b_1} depende de la dispersión de los datos con respecto a la recta de regresión (s_e) y del tamaño de la muestra (n). La fórmula de ET_{b_1} fundamenta esta intuición mostrando que los datos con menor dispersión con respecto a la recta de regresión (s_e más pequeña) y tamaños

muestrales mayores (valores de n mayores) producen estimaciones más precisas de β_1 (es decir, un ET_{b_1} más pequeño). Aunque la variabilidad en Y y el tamaño de la muestra son los únicos dos factores que afectan a la capacidad de estimar de forma precisa una media poblacional (ET_Y), hay un tercer factor que es importante para la estimación precisa de β_1 : la variabilidad de los datos en X . Cuanto más dispersos estén nuestros valores en X (s_x mayor), más precisa será nuestra estimación de β_1 . La dependencia de la dispersión de los valores de X se ilustra en la Figura 12.5.1, que muestra dos conjuntos de datos con el mismo valor de s_e y el mismo valor de n , pero valores diferentes de s_x . Imaginemos que usamos una regla para ajustar una recta a ojo. Es claramente intuitivo que los datos en el caso (b), que tienen un valor de s_x mayor, determinarían la pendiente de la recta de forma más precisa.

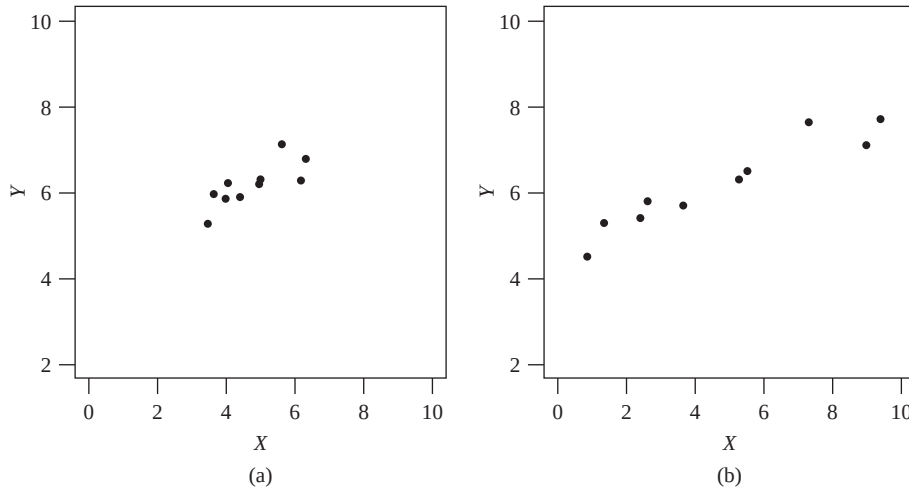


Figura 12.5.1 Dos conjuntos de datos con los mismos valores de n y de s_e pero con s_x diferente: (a) s_x menor y (b) s_x mayor

Como otra forma de plantear estas ideas, imaginemos que extendemos nuestros brazos frente a nosotros, extendemos el dedo índice de cada mano, y equilibramos una regla de medir entre los dos dedos. Si separamos nuestras manos entre sí, equilibrar la regla es fácil [es como el caso (b)]. Sin embargo, si acercamos nuestras manos, equilibrar la regla se hace más difícil [es como el caso (a)]. Aumentando la base de soporte se aumenta la estabilidad. Asimismo, tener los valores de x en un intervalo más amplio disminuye el error típico de la pendiente.

Implicaciones en el diseño. La exposición anterior implica que, para el propósito de obtener información precisa sobre β_1 , es mejor tener los valores de X tan dispersos como sea posible. Este hecho puede servir de guía al investigador cuando el diseño del experimento incluye la elección de valores de X , sin embargo, otros factores juegan también un papel. Por ejemplo, si X es la dosis de un medicamento, el criterio de valores de X ampliamente dispersos podría llevar a utilizar únicamente dos dosis, una muy baja y otra muy alta. Pero en la práctica el investigador desearía tener al menos unas cuantas observaciones de dosis intermedias, para verificar que la relación es realmente lineal dentro del intervalo de los datos.

Intervalo de confianza para β_1

En muchos estudios el valor β_1 es un parámetro con sentido biológico y el objetivo principal del análisis de datos es estimarlo. Se puede construir un intervalo de confianza para β_1 por el método habitual basado en el ET y la distribución t de Student. Por ejemplo

$$b_1 \pm t_{0,025} ET_{b_1}$$

donde el valor crítico $t_{0,025}$ se determina de la distribución t de Student con

$$gl = n - 2$$

Los intervalos con otros coeficientes de confianza se construyen de forma análoga, como por ejemplo para un intervalo de confianza del 90 % utilizaríamos $t_{0,05}$.

Ejemplo 12.5.2 Longitud y peso de serpientes

Utilicemos los datos de serpientes para construir un intervalo de confianza del 95 % para β_1 . Encontramos que $b_1 = 7,19186$ y $ET_{b_1} = 0,9531$. Hay $n = 9$ observaciones. Buscando en la Tabla 4 con $gl = 9 - 2 = 7$, obtenemos

$$t_{7, 0,025} = 2,365$$

El intervalo de confianza es

$$7,19186 \pm 2,365 \times 0,9531$$

o

$$4,94 \text{ g/cm} < \beta_1 < 9,45 \text{ g/cm}$$

Tenemos una confianza del 95 % en que la verdadera pendiente de la regresión del peso con respecto a la longitud para esta población de serpientes esté entre 4,94 g/cm y 9,45 g/cm. Es un intervalo más bien amplio porque el tamaño muestral no es muy grande. ■

Contraste de la hipótesis $H_0: \beta_1 = 0$

En algunas investigaciones no se sabe a priori si existe una relación lineal entre X e Y . En esos casos, puede ser relevante considerar la posibilidad de que cualquier tendencia aparente de los datos es ilusoria y refleja solamente la variabilidad del muestreo. En esta situación es natural formular la hipótesis nula

$$H_0: \mu_{Y|X} \text{ no depende de } X$$

En el modelo lineal, esta hipótesis se puede traducir como

$$H_0: \beta_1 = 0$$

Un test t de H_0 se basa en el estadístico de contraste*

$$t_s = \frac{b_1 - 0}{ET_{b_1}}$$

Los valores críticos se obtienen de la distribución t de Student con

$$gl = n - 2$$

El ejemplo siguiente ilustra la aplicación de este test t .

Ejemplo 12.5.3 Presión sanguínea y calcio en plaquetas

Los datos sobre presión sanguínea y calcio en plaquetas del Ejemplo 12.2.3 se muestra en la Figura 12.5.2. Los cálculos a partir de los datos producen los siguientes resultados: $\bar{x} = 94,50000$, $\bar{y} = 107,86840$, $s_x = 8,04968$, $s_y = 16,07780$, de donde podemos calcular†

$$b_0 = -2,2009 \quad y \quad b_1 = 1,16475$$

La suma de cuadrados residual es 6.311,7618.

Por tanto,

$$s_e = \sqrt{\frac{6.311,76}{38 - 2}} = 13,24 \quad y \quad ET_{b_1} = \frac{13,24}{8,04968 \sqrt{38 - 1}} = 0,2704$$

* Incluimos «-0» en el numerador del estadístico de contraste para recordarnos que estamos comparando nuestra pendiente estimada (observada), b_1 , con la pendiente que esperaríamos observar si la hipótesis nula fuera cierta. En los ejercicios consideraremos una situación en la que la pendiente en la hipótesis nula puede tomar un valor distinto de cero.

† Como los valores siguientes son cálculos intermedios utilizados en la regresión, incluimos más dígitos que los que se presentarían generalmente en un informe.

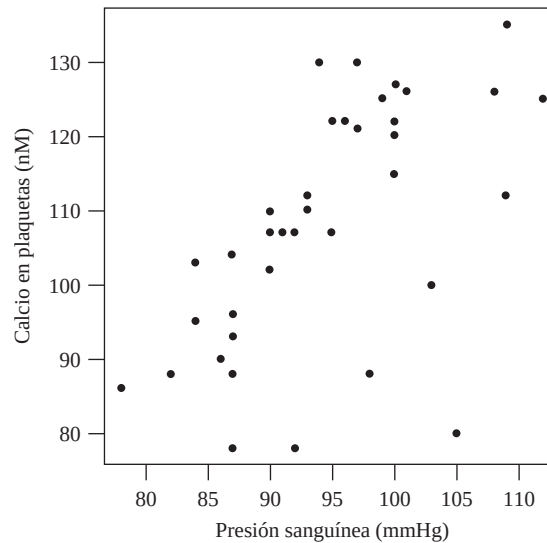


Figura 12.5.2 Presión sanguínea y calcio en plaquetas para 38 personas con presión sanguínea normal

Los valores de b_0 , b_1 , $SC(\text{resid})$ y ET_{b_1} se obtienen generalmente utilizando software estadístico. Mostramos a continuación una salida típica de computador:

```
The regression equations is
Calcio en Plaquetas=-2.2 +1.16 Presión Sanguínea
Predictor      Coef    SE Coef      T      P
Constant       -2.20    25.65     -0.09  0.932
Presión Sanguínea 1.1648    0.2704     4.31  0.000
S=13.2411 R-Sq=34.0% R-Sq(adj)=32.2%

Analysis of Variance
Source      DF      SS      MS      F      P
Regression    1  3252.6  3252.6  18.55  0.000
Residual Error 36  6311.8  175.3
Total        37  9564.3
```

Contrastaremos la hipótesis nula

$$H_0: \beta_1 = 0$$

frente a la alternativa no direccional

$$H_A: \beta_1 \neq 0$$

Estas hipótesis son traducciones, en el modelo lineal, de las hipótesis verbales

H_0 : El calcio medio en plaquetas no está relacionado linealmente con la presión sanguínea

H_A : El calcio en plaquetas está relacionado linealmente con la presión sanguínea

(Nota: «Relacionado linealmente» no se refiere necesariamente a una dependencia causal, como ya hemos comentado en la Sección 12.2).

Escojamos $\alpha = 0,05$. El estadístico de contraste es

$$t_s = \frac{1,16475}{0,2704} = 4,308$$

Utilizando la Tabla 4 con $gl = n - 2 = 36 \approx 40$, obtenemos $t_{40, 0,0005} = 3,551$. Por tanto, resulta P valor $< 0,001$ y rechazamos H_0 . Los datos muestran evidencia suficiente (y muy fuerte) para concluir que la verdadera pendiente de la recta de regresión del calcio en plaquetas con respecto a la presión sanguínea en esta población es positiva (es decir, $\beta_1 > 0$). ■

Nótese que el contraste sobre β_1 no pregunta si la relación entre $\mu_{Y|X}$ y X es lineal. Más bien, el test pregunta si, *asumiendo* que el modelo lineal es aplicable, podemos concluir que la pendiente no es cero. Por tanto, es necesario ser cuidadosos al plantear la conclusión de este test. Por ejemplo, la afirmación «hay una tendencia lineal significativa», podría ser fácilmente malinterpretada*.

Como en el caso de otros contrastes de hipótesis, si deseamos utilizar una hipótesis alternativa direccional, seguiremos el procedimiento en dos pasos de (1) comprobar si la dirección especificada es correcta (que en un escenario de regresión significa comprobar si la pendiente de la recta de regresión tiene el signo $+$ o $-$ correcto) y (2) dividir por dos el P valor no direccional si la condición anterior se cumple.

Ejercicios 12.5.1-12.5.9

12.5.1 En referencia a los datos sobre la leucina dados en el Ejercicio 12.3.1:

- Construya un intervalo de confianza del 95 % para β_1 .
- Interprete el intervalo de confianza del apartado (a) en el contexto de este ejercicio.

12.5.2 En referencia a datos sobre temperatura corporal del Ejercicio 12.3.2, para esos datos, $s_e = 0,91472$. Construya un intervalo de confianza del 95 % para β_1 .

12.5.3 En referencia a los datos sobre peso de mazorcas del Ejercicio 12.2.5, para esos datos, $SC(\text{resid}) = 1.337,3$.

- Construya un intervalo de confianza del 95 % para β_1 .
- Interprete el intervalo de confianza del apartado (a) en el contexto de este ejercicio.

12.5.4 En referencia a los datos sobre crecimiento de hongos del Ejercicio 12.2.6, para esos datos, $SC(\text{resid}) = 16,7812$.

- Calcule el error típico de la pendiente, ET_{b_1} .
- Considere la hipótesis nula de que el ácido laetisárico no tiene efecto sobre el crecimiento de los hongos. Suponiendo que el modelo lineal es aplicable, fórmulela como una hipótesis sobre la verdadera recta de regresión, y contraste la hipótesis frente a la alternativa de que el ácido laetisárico inhibe el crecimiento de los hongos. Utilice $\alpha = 0,05$.

12.5.5 En referencia a los datos sobre gasto de energía del Ejercicio 12.2.7, para esos datos, $SC(\text{resid}) = 21.026,1$.

- Construya un intervalo de confianza del 95 % para β_1 .
- Construya un intervalo de confianza del 90 % para β_1 .

12.5.6 En referencia a los datos sobre bomba de Ca basal del Ejercicio 12.2.8, para estos datos, $s_e = 548,78$.

- Construya un intervalo de confianza del 95 % para β_1 .

- ¿Qué piensa sobre la afirmación de que β_1 es menor que -800 (nmol/mg/h)/(μg/ml)? Utilice el intervalo del apartado (a) para fundamentar su respuesta.

- ¿Qué piensa sobre la afirmación de que β_1 es menor que 800 (nmol/mg/h)/(μg/ml) en valor absoluto? Utilice el intervalo del apartado (a) para fundamentar su respuesta.

12.5.7 En referencia a los datos sobre respiración del Ejercicio 12.3.6, suponiendo que el modelo lineal es aplicable, contraste la hipótesis nula de no relación frente a la alternativa de que los árboles de altitudes mayores tienden a tener tasas de respiración mayores. Utilice $\alpha = 0,05$.

12.5.8 La siguiente salida de computador es de un ajuste de un modelo de regresión a los datos sobre longitud de serpientes del Ejemplo 12.2.2. Utilice esta salida para construir un intervalo de confianza del 95 % para β_1 .

```
The regression equation is
Peso=-301+7.19Longitud
Predictor    Coef    Stdev    t-ratio    p
Constant   -301.09    60.19     -5.00     0.000
Longitud     7.1919    0.9531     7.55     0.000

S=12.50  R-Sq=89.1%  R-Sq(adj)=87.5%

Analysis of Variance
SOURCE      DF      SS      MS      F      p
Regression    1  8896.3  8896.3  56.94  0.000
Error         7  1093.7   156.2
Total         8  9990.0
```

12.5.9 En referencia a los datos sobre flujo de pico del Ejercicio 12.3.8, suponga que el modelo lineal es aplicable.

- Contraste la hipótesis nula de no relación frente a la alternativa de que el flujo de pico está relacionado con la altura. Utilice una alternativa no direccional y $\alpha = 0,10$.
- Repita el test del apartado (a), pero esta vez utilice la alternativa direccional de que el flujo de pico tiende a aumentar con la altura. Use de nuevo $\alpha = 0,10$.

* Hay tests que pueden (en algunas circunstancias) contrastar si la verdadera relación es lineal. Además, existen tests que pueden contrastar la existencia de una componente lineal o tendencia sin suponer que la relación es lineal. Estos tests están fuera del alcance de este libro.

12.6 Pautas para la interpretación de la regresión y de la correlación

Cualquier conjunto de datos (X, Y) se puede someter a un análisis de regresión y se pueden calcular los valores de b_0 , b_1 , s_e y r . Pero esos valores deben ser interpretados con cuidado. En esta sección presentamos pautas y precauciones para la interpretación de la regresión lineal y de la correlación. Consideraremos en primer lugar el uso de la regresión y la correlación para propósitos puramente descriptivos y después pasaremos a los usos para inferencia.

¿Cuándo es descriptivamente inadecuada la regresión lineal?

La regresión lineal y la correlación pueden proporcionar una descripción inadecuada de un conjunto de datos si está presente cualquiera de las siguientes características:

- Carácter curvilíneo.
- *Outliers*.
- Puntos de influencia.

Presentaremos brevemente cada una de ellas.

Si la dependencia de Y con respecto a X es realmente curvilínea en vez de lineal, la aplicación de la regresión lineal y la correlación puede ser muy engañosa. El ejemplo siguiente muestra lo que puede ocurrir.

Ejemplo 12.6.1 Una relación curvilínea con respecto a X

La Figura 12.6.1 muestra un conjunto de datos ficticios que sigue la relación exacta: $Y = -1 + 6X - X^2$. No obstante, X e Y están incorrelacionados: $r = 0$ y $b_1 = 0$. La mejor línea recta que se ajusta a esos datos sería una recta horizontal, pero, por supuesto, dicha recta sería un resumen muy pobre de la relación curvilínea entre X e Y . La DT residual es $s_e = 2,27$. Sin embargo, como estos datos no son aleatorios, no mide variación aleatoria, sino que más bien mide la desviación de la linealidad. ■

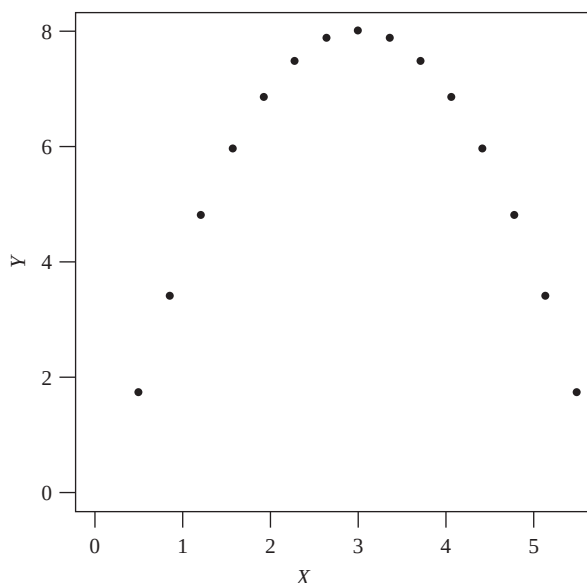


Figura 12.6.1 Datos para los que X e Y están incorrelacionados pero tienen una fuerte relación curvilínea

Generalmente, las consecuencias de las relaciones curvilíneas son que (1) la recta ajustada no representa adecuadamente a los datos; (2) la correlación es engañosamente pequeña; (3) s_e está aumentada. Por supuesto, el Ejemplo 12.6.1 es un caso extremo de esta distorsión. La Figura 12.6.2 muestra un conjunto de datos con una relación curvilínea débil, pero que todavía puede notarse.

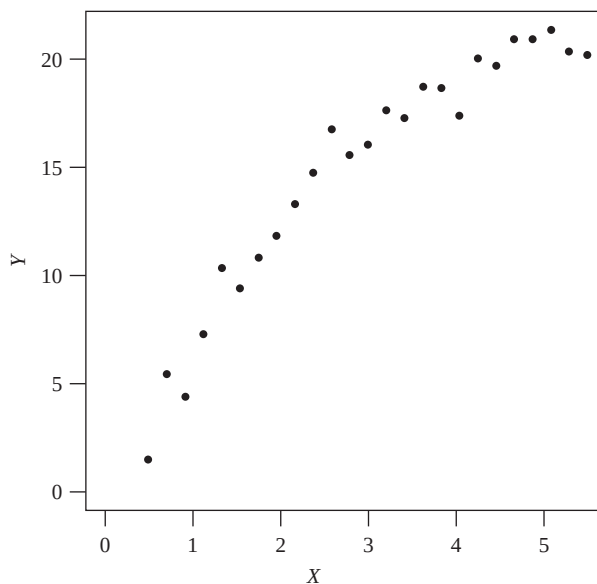


Figura 12.6.2 Datos que muestran una relación curvilínea débil

Los **outliers** en un escenario de regresión son puntos que están inusualmente lejos de la tendencia lineal formada por los datos. Los **outliers** pueden distorsionar el análisis de regresión de dos formas: (1) aumentando s_e y reduciendo la correlación y (2) influye excesivamente en la recta de regresión. Nótese que un punto puede ser un **outlier** en un diagrama de dispersión sin serlo en los valores de la distribución de X o en los valores de la distribución de Y , como veremos en el ejemplo siguiente.

La Figura 12.6.3 muestra un conjunto de datos con una variedad de **outliers**. La Figura 12.6.3(a) muestra un conjunto de datos sin **outliers**, mientras que (b) y (c) muestran datos con **outliers** de regresión, es decir, con puntos que están lejos de la recta de regresión. En la gráfica (b) el punto **outlier** no parece afectar mucho a la recta de regresión, pero aumenta la desviación típica residual s_e y reduce la correlación. El punto **outlier** de la gráfica (c) parece afectar grandemente a la pendiente de la recta de regresión estimada. Además, aumenta también s_e y reduce la correlación. Aunque el punto inusual en (d) es un **outlier** con respecto a la distribución X (y a la Y), no es un **outlier** en el contexto de la regresión ya que no está situado lejos de la recta de regresión.

Los **puntos de apalancamiento** son puntos que tienen el potencial de influir grandemente en la pendiente del modelo de regresión ajustado. Cuanto más lejos esté el valor X de un punto del centro de la distribución X , más influencia tiene ese punto en el modelo de regresión global. *Tener influencia y ejercerla* realmente son sin embargo dos cosas diferentes. La Figura 12.6.3(c) y (d) muestran ejemplos de puntos de apalancamiento. En la gráfica (c), el punto de apalancamiento ejerce realmente su influencia en la recta inclinando la regresión y desviándola del grueso de los datos. Un punto que tiene un gran efecto en el modelo de regresión se denomina **punto de influencia**. La gráfica (d) muestra un punto de apalancamiento (debido al valor extremo de la X) que no es de influencia porque la recta de regresión no se desvía de la tendencia del grueso de los datos. Nótese que el **outlier** de la gráfica (b) no se considera un punto de apalancamiento, ya que su capacidad para desviar la pendiente de la línea es escasa debido a que su valor en X está cerca del centro de la distribución.

Los puntos de influencia pueden afectar también grandemente (aumentando o disminuyendo) el tamaño del coeficiente de correlación. En la Figura 12.6.3, el punto de influencia de (c) disminuye la correlación desde un valor de $r = 0,956$ en (a) hasta $r = 0,579$. El Ejemplo 12.6.3 muestra una situación en la que la correlación aumenta por la presencia de un punto de influencia.

La Figura 12.6.4(a) muestra un conjunto de datos y una recta de regresión. La Figura 12.6.4(b) muestra el mismo conjunto de datos, pero en el que se ha añadido un punto de influencia. Al incluir dicho punto de influencia, en el conjunto de datos, la recta de regresión cambia de forma apreciable. Aunque el punto de influencia es un **outlier** en las distribuciones X e Y , no es un **outlier** de la regresión ya que el residuo para ese punto no es muy grande.

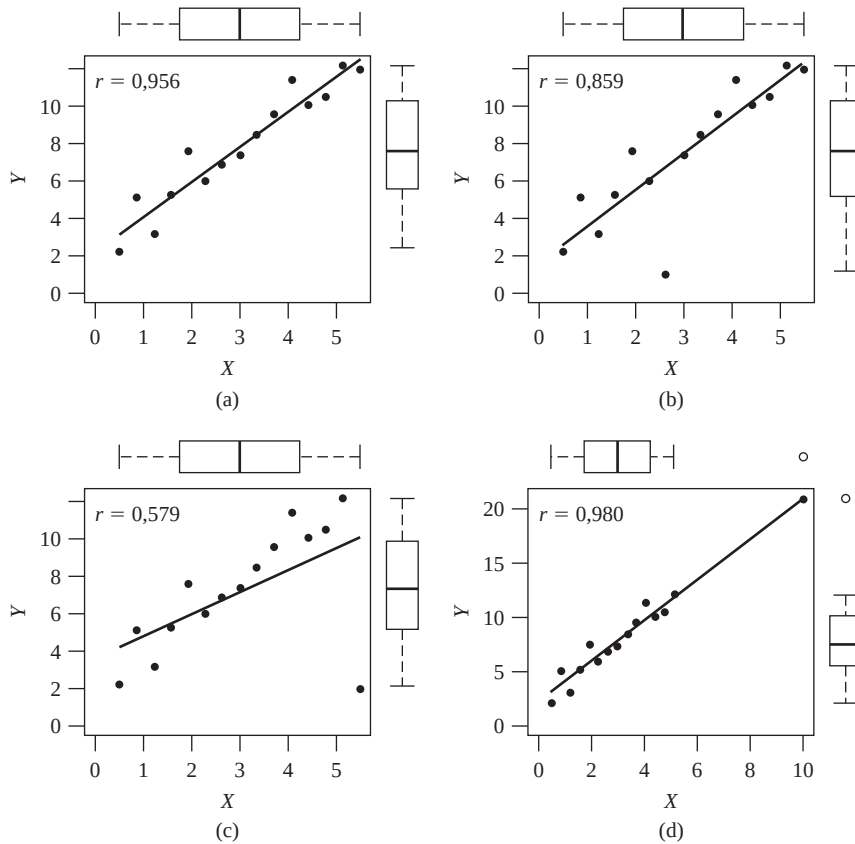


Figura 12.6.3 Efectos diferentes de *outliers* en la recta de regresión. Los diagramas de X e Y aparecen en los márgenes de cada diagrama de dispersión. (a) Un conjunto de datos sin *outliers*; (b) Los mismos datos, excepto porque hay un *outlier* en la parte media de los valores de X; (c) Los mismos datos, excepto porque hay un *outlier* en el extremo superior de los valores de X (un punto de apalancamiento e influencia); (d) Los mismos datos, excepto porque hay un *outlier* con respecto a la distribución X (e Y), pero no con respecto a la recta de regresión (un punto del apalancamiento, pero con poca influencia)

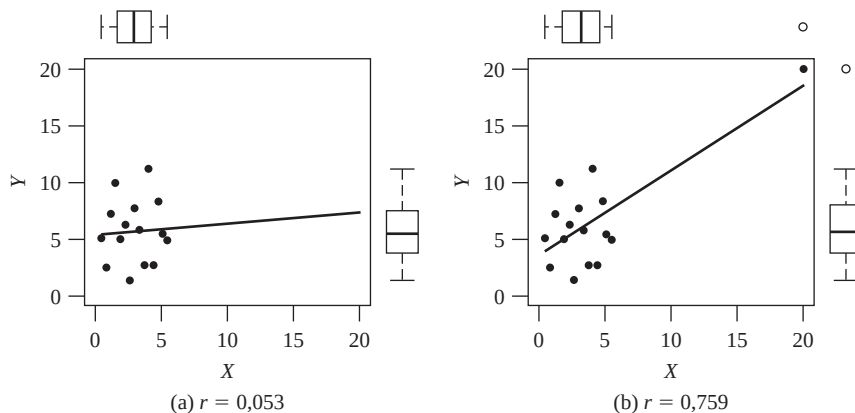


Figura 12.6.4 Efecto de un punto de influencia en la recta de regresión. (a) Un conjunto de datos; (b) Los mismos datos con un punto de influencia añadido

El coeficiente de correlación de los datos de la Figura 12.6.4(a) es $r = 0,053$. Al añadir el punto de influencia al conjunto de datos, la correlación cambia a $r = 0,759$, para los datos de la Figura 12.6.4(b).

Condiciones para la Inferencia

Los valores b_0 , b_1 , s_e y r se pueden utilizar para describir un diagrama de dispersión que muestre una tendencia lineal. Sin embargo, la inferencia estadística basada en estos valores depende de ciertas condiciones relativas al diseño del estudio, a los parámetros y a las distribuciones poblacionales condicionales. Resumimos estas condiciones y después expone-mos pautas y precauciones con respecto a las mismas.

1. *Condiciones de diseño.* Hemos presentado dos modelos de muestreo para la regresión y la correlación:
 - (a) Modelo de submuestreo aleatorio: Para cada X observado, el correspondiente valor de Y se considera elegido aleatoriamente de la distribución poblacional condicional de los valores de Y para ese X^* .
 - (b) Modelo de muestreo aleatorio bivariado: Cada pareja observada (X, Y) se considera elegida aleatoriamente de la distribución poblacional conjunta de parejas bivariadas (X, Y) .
 En cualquiera de los dos modelos de muestreo, cada pareja observada (X, Y) debe ser independiente de las otras. Esto significa que el diseño experimental no debe incluir emparejamiento, bloque ni estructuras jerárquicas.
2. *Condiciones con respecto a los parámetros.* El modelo lineal establece que
 - (a) $\mu_{Y|X} = \beta_0 + \beta_1 X$.
 - (b) σ_e no depende de X .
3. *Condición con respecto a las distribuciones poblacionales.* El intervalo de confianza y el test t se basan en que la distribución poblacional condicional de Y para cada valor fijo de X es normal.

El modelo de submuestreo aleatorio se requiere si b_0 , b_1 y s_e se ven como estimadores de los parámetros β_0 , β_1 y σ_e , y mencionados en el modelo lineal. El modelo de muestreo aleatorio bivariado se requiere si r se ve como estimador del parámetro poblacional ρ . Se puede demostrar que si el modelo de muestreo aleatorio bivariado es aplicable, entonces también es aplicable el modelo de submuestreo aleatorio. Por tanto, los parámetros de la regresión siempre se pueden estimar si la correlación se puede estimar, pero no al revés.

Pautas con respecto a las condiciones de muestreo

Apartarse de las condiciones de muestreo no solo afecta a la validez de las técnicas normales como el intervalo de confianza para β_1 , sino que puede conducir a una interpretación falsa de los datos incluso aunque no se realice ningún análisis estadístico formal. Dos errores de interpretación que aparecen algunas veces en la práctica son: (1) fallo al tener en cuenta la dependencia de las observaciones y (2) precaución insuficiente al interpretar r cuando los valores de X no representan una muestra aleatoria.

Los dos ejemplos siguientes ilustran estudios con observaciones independientes.

Ejemplo 12.6.2 Colesterol en suero y glucosa en suero

Supongamos un conjunto de datos formado por 20 parejas de medidas de colesterol en suero (X) y de glucosa en suero (Y) en humanos. Sin embargo el experimento incluyó solo dos sujetos, cada uno de ellos medido en 10 ocasiones diferentes. Debido a la dependencia de los datos, no es correcto tratar inocentemente los 20 puntos del mismo modo. La Figura 12.6.5 ilustra la dificultad. Dicha figura muestra que no hay evidencia de correlación alguna entre X e Y , excepto por el modesto hecho de que el sujeto con valores mayores de X tiene también valores mayores de Y . Claramente, sería imposible interpretar apropiadamente el diagrama de dispersión si los 20 puntos se hubieran representado con el mismo símbolo. Por el mismo motivo, la aplicación de las fórmulas de regresión o de correlación a las 20 observaciones sería seriamente erróneo²⁰. ■

* Si la variable X incluye error de medida, entonces X se debe interpretar en el modelo lineal como el valor medido de X , en vez de como algún valor subyacente «verdadero» de X . Los modelos lineales que consideran el «verdadero» valor de X conducen a otras clases de análisis de regresión.

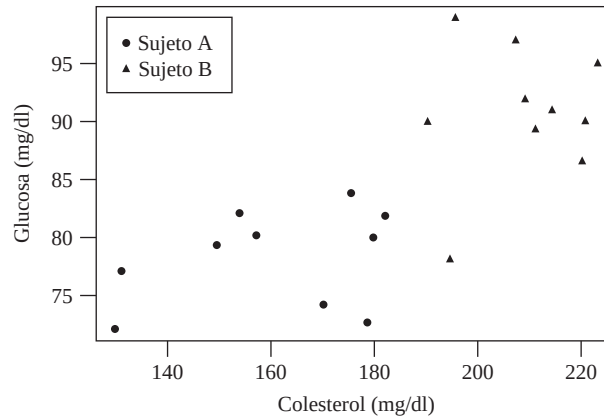


Figura 12.6.5 Veinte observaciones de X = colesterol en suero e Y = glucosa en suero en humanos

Ejemplo 12.6.3 Crecimiento de novillos

La Figura 12.6.6 muestra 20 parejas de medidas de peso (Y) de novillos en varios momentos (X) durante un experimento sobre alimentación. Los datos representan cuatro animales, y cada uno se pesó cinco veces. Las observaciones del mismo animal están unidos por líneas en la figura. Un análisis de regresión ordinario con los 20 puntos de datos ignoraría la información que llevan las líneas y produciría valores de ET aumentados y tests débiles. De forma similar, un diagrama de dispersión ordinario (sin las líneas) sería una representación inadecuada de los datos²¹.

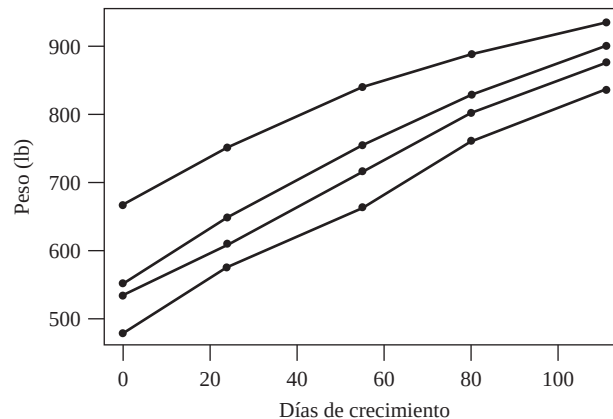


Figura 12.6.6 Veinte observaciones de X = días e Y = peso en novillos. Los datos de cada animal están unidos por líneas

En el Ejemplo 12.6.2, ignorar la dependencia en las observaciones conduciría a una *sobreinterpretación* de los datos (es decir, concluir que hay relación cuando en realidad hay poca evidencia a su favor). Por el contrario, ignorar la dependencia en el Ejemplo 12.6.3 conduciría a una *subinterpretación* de los datos (es decir, una extracción insuficiente de la «señal» en el «ruido»).

Al interpretar el coeficiente de correlación r , hay que tener en cuenta que dicho coeficiente está influido por el grado de dispersión de los valores de X . Si los valores de la regresión b_0 , b_1 , y s_e no cambian, *cuanto más dispersos estén los valores de X , mayor correlación (mayor valor absoluto de r)*. El ejemplo siguiente muestra cómo ocurre esto.

Ejemplo 12.6.4

La Figura 12.6.7 muestra datos ficticios que ilustran cómo resulta afectado r por la distribución de X . Los puntos de los datos de las gráficas (a) y (b) se han representado juntos en la gráfica (c). La recta de regresión es casi la misma en los tres diagramas de dispersión, pero nótese que X e Y aparecen más altamente correlacionados en (c) que en (a) o (b). La apariencia de los diagramas de dispersión se refleja en los coeficientes de correlación. De hecho, $r = 0,60$ en (a), $r = 0,58$ en (b), pero $r = 0,85$ en (c).

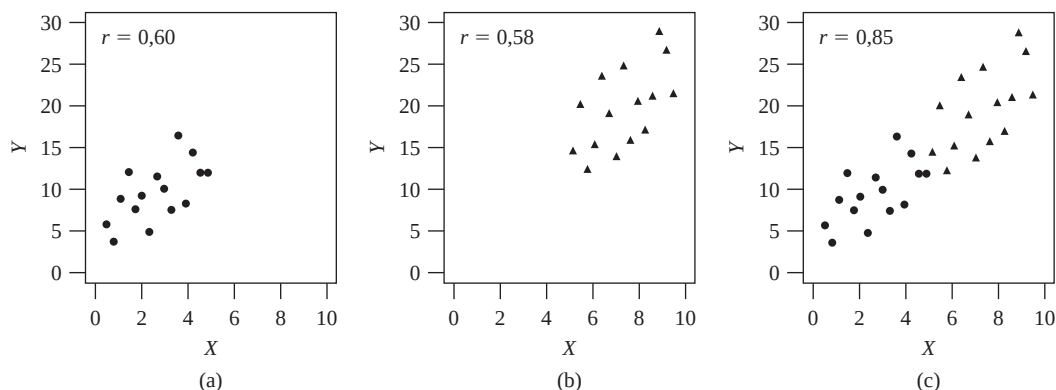


Figura 12.6.7 Dependencia de r en la distribución de X . Los datos de (a) y (b) se representan juntos en (c)

El hecho de que r dependa de la distribución de X no significa que r no sea válido como estadístico descriptivo. Pero significa que, cuando los valores de X no se puede considerar como una muestra aleatoria, r debe ser interpretado con cuidado. Por ejemplo, supongamos que dos experimentadores realizan estudios separados de respuesta (Y) a varias dosis (X) de un medicamento. Cada uno de ellos calcula r como una descripción de sus propios datos, pero *no* deberían esperar obtener valores *similares* de r a menos que utilicen las mismas dosis (valores de X). Por el contrario, se podría esperar razonablemente obtener rectas de regresión similares y desviaciones típicas residuales similares, independientemente de la elección de los valores de X , siempre que la relación dosis-respuesta sea la misma en el intervalo de dosis utilizado.

Etiquetado de X e Y Si es aplicable el modelo de muestreo aleatorio bivariado, entonces el investigador es libre de decidir a qué variable denominar X y a qué variable denominar Y . Por supuesto, para el cálculo de r el etiquetado no importa. Para los cálculos de regresión, la decisión depende del propósito del análisis. La regresión de Y con respecto a X produce (en el modelo lineal) estimadores de $\mu_{Y|X}$, es decir, de la media poblacional de los valores de Y para un X fijo. De forma similar, la regresión de X con respecto a Y está dirigida a estimar $\mu_{X|Y}$, es decir, el valor medio de X para un Y fijo. Estos planteamientos no conducen a la misma recta de regresión porque están dirigidos a responder a preguntas diferentes. Sigue a continuación un ejemplo intuitivo.

Ejemplo 12.6.5 Altura y peso de hombres jóvenes

Para la población de hombres jóvenes descrita en el Ejemplo 12.4.4, el peso medio de hombres jóvenes cuya altura es 76'' es de 178 lb. Consideremos ahora esta cuestión: ¿Cuál sería la altura media de los hombres jóvenes que pesan 178 lb? No hay razón para que la respuesta sea 76''. La intuición sugiere que la respuesta debería ser menor que 76'', y de hecho es aproximadamente 71''.

Pautas con respecto al modelo lineal y a la condición de normalidad

El test y el intervalo de confianza para β_1 se basan en el modelo lineal y la condición de normalidad. La interpretación de esas inferencias se puede degradar seriamente si la condición de linealidad no se cumple. Después de todo, hemos visto anteriormente en esta sección que incluso el uso descriptivo de la regresión queda reducido si están presentes comportamientos curvilíneos o *outliers*.

Además de la linealidad, el modelo lineal especifica que σ_e es la misma para todas las observaciones. Una forma común de apartarse de esta condición es una tendencia a que las medias mayores estén asociadas con DT mayores. Un valor ligeramente no constante de la DT no afecta seriamente a la interpretación de b_0 , b_1 , ET_{b_1} y r (aunque invalida la interpretación de s_e como estimador conjunto de una DT común).

Gráficas de residuos

Los tests estadísticos para la condición curvilínea, desviaciones típicas no iguales, no normalidad y *outliers* están fuera del alcance de este libro. Sin embargo, el instrumento más útil para detectar esas características es el ojo humano, ayudado por los diagramas de dispersión. Por ejemplo, nótese lo fácilmente que el ojo detecta la débil curvatura de la Figura 12.6.2 y el *outlier* de la Figura 12.6.3(b). Nótese también que en la Figura 12.6.3(b), el examen separado de las distribuciones marginales de X e Y no habría revelado el *outlier*.

Además de los diagramas de dispersión de Y con respecto a X , a menudo es útil observar varias gráficas de los residuos. El diagrama de dispersión de los residuos ($y_i - \hat{y}_i$) en función de $\hat{y}_i$ se denomina **gráfica de residuos**. Las gráficas de residuos son muy útiles para detectar curvaturas. Pueden revelar también tendencias en la desviación típica condicional. La Figura 12.6.8 muestra los datos de la Figura 12.6.2 junto con una gráfica de residuos de dichos datos.

La gráfica de residuos muestra los datos tras eliminar la tendencia lineal, lo que facilita ver si los datos muestran patrones no lineales. La curvatura de la Figura 12.6.8(a) es aparente, pero resulta mucho más visible en la gráfica de residuos de la Figura 12.6.8(b).

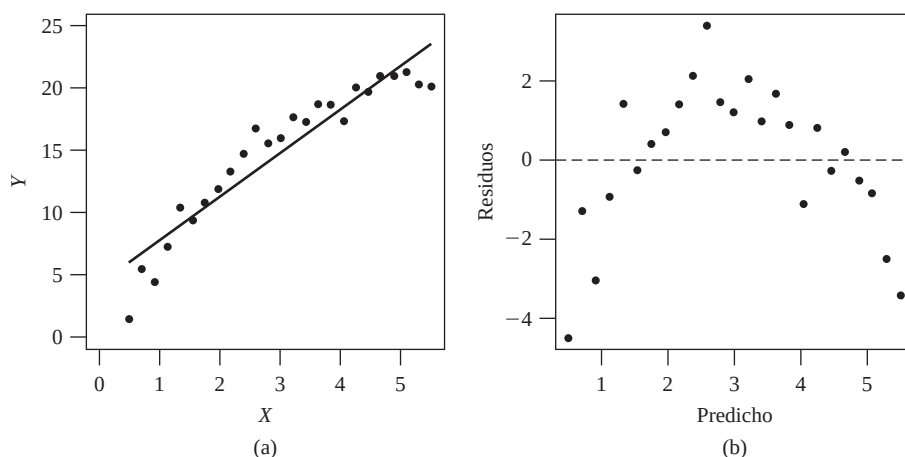


Figura 12.6.8 Datos que muestran una curvatura leve con una recta de regresión lineal; (b) gráfica de residuos de los datos

Si el modelo lineal es aplicable, y no hay *outliers*, entonces la recta de regresión ajustada recoge la tendencia de los datos, dejando un patrón aleatorio en la gráfica de residuos. Por tanto, *no esperamos ver ningún patrón llamativo en una gráfica de residuos*. Por ejemplo, la Figura 12.6.9 muestra una gráfica de residuos de los datos sobre serpientes del Ejemplo 12.2.1. La falta de características inusuales de esta gráfica fundamenta el uso de un modelo de regresión para estos datos.

Si se cumple la condición de normalidad, entonces la distribución de los residuos tendría que tener el aspecto de una distribución normal*. La gráfica de probabilidad normal de los residuos es una comprobación útil de la condición de normalidad. La gráfica de probabilidad de los datos sobre serpientes de la Figura 12.6.10 es bastante lineal, lo que fundamenta el uso del test t y el intervalo de confianza presentados en la Sección 12.5.

* Esta es la base de las interpretaciones del 68 % y del 95 % de s_e dadas en la Sección 12.3.

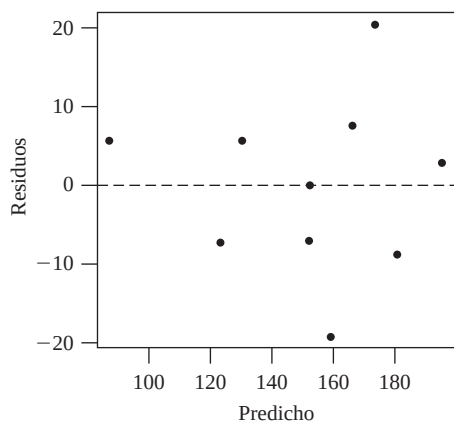


Figura 12.6.9 Gráfica de residuos de los datos sobre serpientes

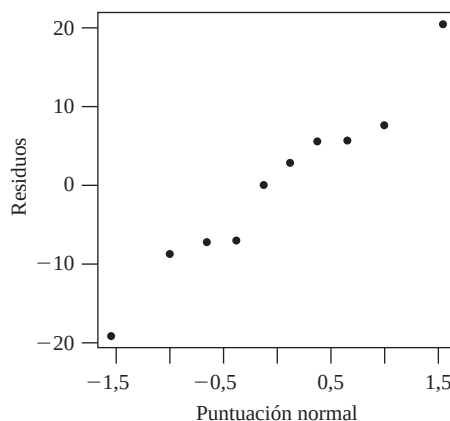


Figura 12.6.10 Gráfica de probabilidad normal de los datos sobre serpientes

Uso de transformaciones

Si las condiciones de linealidad, desviación típica constante y normalidad no se cumplen, un remedio que a veces es útil es transformar la escala de medida de Y , X o ambas. El ejemplo siguiente ilustra el uso de una transformación logarítmica.

Ejemplo 12.6.6 Crecimiento de plantas de soja

Una botánica plantó 60 plantas de soja de una semana de edad en tiestos individuales. Tras 12 días de crecimiento, cosechó, secó y pesó 12 de las plantas de soja jóvenes. Pesó otras 12 plantas tras 23 días de crecimiento, y grupos de 12 plantas tras 27 días, 31 días y 34 días. La Figura 12.6.11 muestra los 60 pesos de las plantas en función de los días de crecimiento. Una curva suave conecta las medias de los grupos. Es fácil ver en la Figura 12.6.11 que la relación entre el peso medio de las plantas y el tiempo es curvilínea en vez de lineal, y que la desviación típica condicional no es constante, sino fuertemente creciente²².

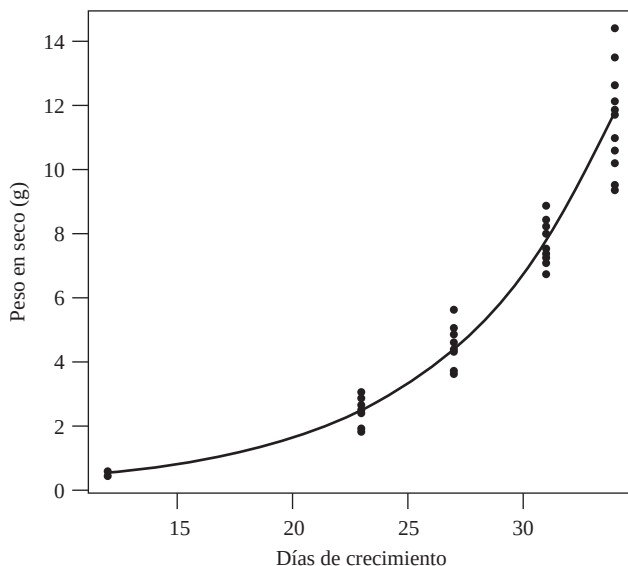


Figura 12.6.11 Peso de las plantas de soja en función de los días de crecimiento

La Figura 12.6.12 muestra los logaritmos (base 10) de los pesos de las plantas, en función de los días de crecimiento, junto con una recta de regresión. Nótese que la transformación logarítmica ha enderezado la curva y prácticamente igualado las desviaciones típicas. No sería irrazonable suponer que el modelo lineal es válido para las variables $Y = \log(\text{peso en seco})$ y $X = \text{días de crecimiento}$. La Tabla 12.6.1 muestra las medias y las desviaciones típicas antes y después de la transformación logarítmica. Nótese especialmente el efecto de la transformación sobre la igualdad de las DT.

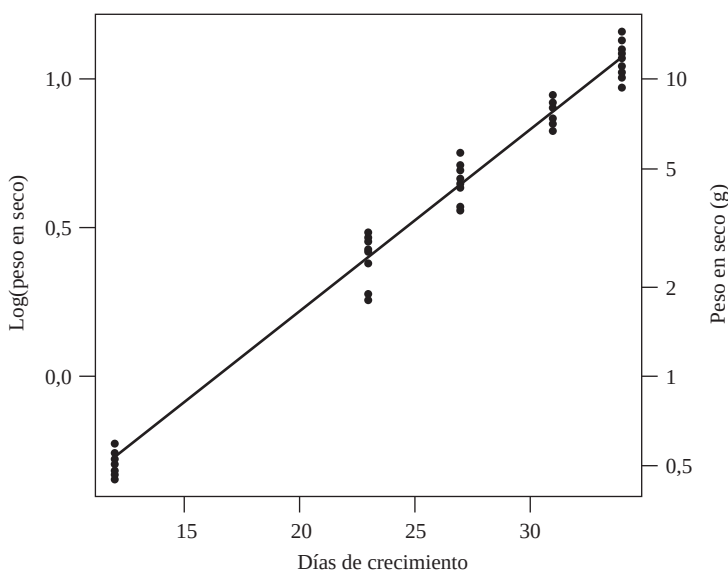


Figura 12.6.12 Log(peso) de las plantas de soja en función de los días de crecimiento

Tabla 12.6.1 Resumen de los datos de crecimiento de plantas de soja en la escala original y después de la transformación logarítmica					
Días de crecimiento	Número de plantas	Peso en seco (g)		Log(peso en seco)	
		Media	DT	Media	DT
12	12	0,50	0,06	-0,31	0,055
23	12	2,63	0,37	0,42	0,062
27	12	4,67	0,70	0,67	0,066
31	12	7,57	1,19	0,87	0,069
34	12	11,20	1,62	1,04	0,064

Ejercicios 12.6.1-12.6.9

12.6.1 En un estudio sobre el metabolismo, se probó a cuatro cerdos macho tres veces: cuando pesaban 30 kg, de nuevo cuando pesaban 60 kg y de nuevo cuando pesaban 90 kg. Durante cada prueba, el investigador analizó la ingestión de alimentos y la salida fecal y urinaria durante 15 días, y con

esos datos calculó el balance de nitrógeno, que se define como la cantidad de nitrógeno incorporado a los tejidos corporales por día. Los resultados se muestran en la tabla siguiente²³.

Número de animal	Balance de nitrógeno (g/día)		
	Peso corporal 30 kg	60 kg	90 kg
1	15,8	21,3	16,5
2	16,4	20,8	18,2
3	17,3	23,8	17,8
4	16,4	22,1	17,5
Media	16,48	22,00	17,50

Suponga que los datos se analizan mediante regresión lineal. Con X = peso corporal e Y = balance de nitrógeno, los cálculos preliminares dan como resultado $\bar{x} = 60$ e $\bar{y} = 18,7$. La pendiente es $b_1 = 0,017$, con un error típico de $ET_{b_1} = 0,032$. El estadístico t es $t_s = 0,53$, que no es significativo con ningún nivel de significación razonable. De acuerdo con este análisis, hay insuficiente evidencia para concluir que el balance de nitrógeno depende del peso corporal, bajo las condiciones de este estudio.

El análisis anterior presenta dos errores. ¿Cuáles son? (*Sugerencia*: busque las formas en las que no se cumplen las condiciones para la inferencia. Hay varios incumplimientos menores de dichas condiciones, pero se pregunta sobre dos importantes. No se requieren cálculos).

12.6.2 Para examinar la digestibilidad de plantas de forraje, se pueden utilizar dos métodos: el material de las plantas se puede fermentar con fluidos digestivos en un contenedor de vidrio, o se puede suministrar a un animal. En cada caso, la digestibilidad se expresa como porcentaje de materia seca total que ha sido digerida. Dos investigadores realizaron estudios separados para comparar los dos métodos, proporcionando varios tipos de forraje a cada método y comparando los resultados. El investigador A obtuvo un valor de correlación de $r = 0,8$ entre los valores de digestibilidad obtenidos por los dos métodos y el investigador B obtuvo $r = 0,3$. La aparente discrepancia entre los resultados quedó resuelta cuando se vio que uno de los investigadores solo había utilizado variedades de pasto cinto (cuya digestibilidad varía entre el 56 % y el 65 %), mientras que el otro investigador había utilizado una variedad de plantas mucho mayor, con digestibilidades que variaban desde el 35 % de los tallos de maíz hasta el 72 % del heno de Fleo²⁴.

¿Qué investigador (A o B) utilizó solo pasto cinto? ¿Cómo explica la diferente selección de los materiales de prueba la discrepancia entre los coeficientes de correlación?

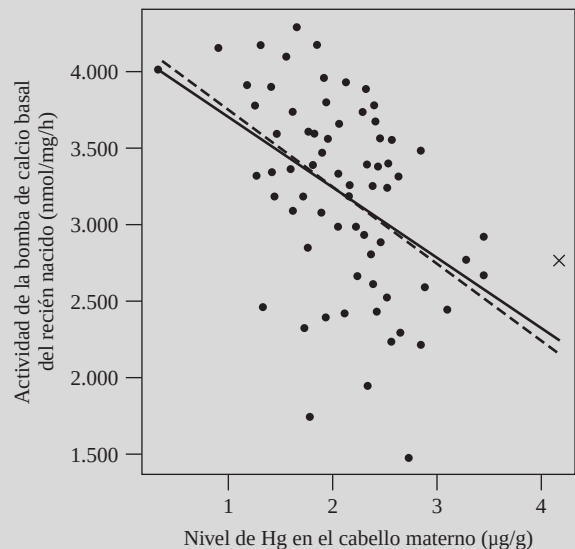
12.6.3 En referencia a los datos de gasto de energía del Ejercicio 12.2.7, el valor de energía de cada sujeto (Y) es el promedio de dos medidas realizadas en diferentes ocasiones. Se podría proponer que sería mejor utilizar las dos medidas como conjuntos de datos separados, con lo que se tendrían 14 observaciones en vez de 7. Si se utilizara este planteamiento

propuesto, una de las condiciones para la inferencia sería altamente dudosa. ¿Cuál y por qué?

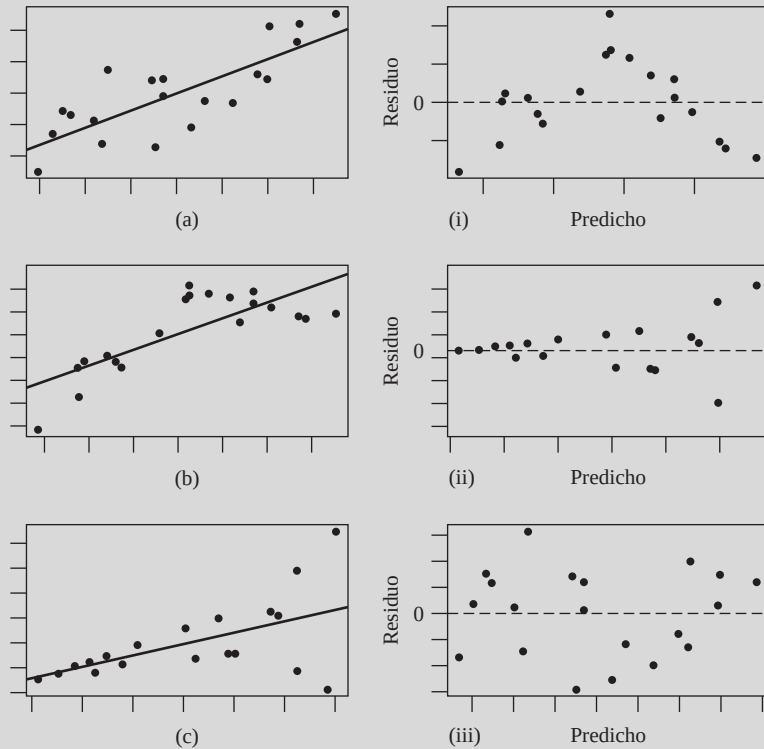
12.6.4 En referencia a los datos sobre crecimiento de hongos del Ejercicio 12.2.6, en dicho ejercicio el investigador obtuvo que $r = -0,98754$. Suponga que un segundo investigador fuera a replicar el experimento, utilizando concentraciones de 0, 2, 4, 6, 8 y 10 mg, con dos placas de Petri en cada concentración. ¿Se podría predecir que el valor de r calculado por este segundo investigador sería aproximadamente el mismo que el obtenido en el Ejercicio 12.2.6, sería de menor valor absoluto o de mayor valor absoluto? Explique su respuesta.

12.6.5 En el siguiente diagrama de dispersión de los datos de la bomba de Ca del Ejercicio 12.2.8, uno de los puntos está marcado con «×». Además, hay dos rectas de regresión en la gráfica. La recta continua incluye todos los datos y la recta discontinua omite el punto marcado como «×».

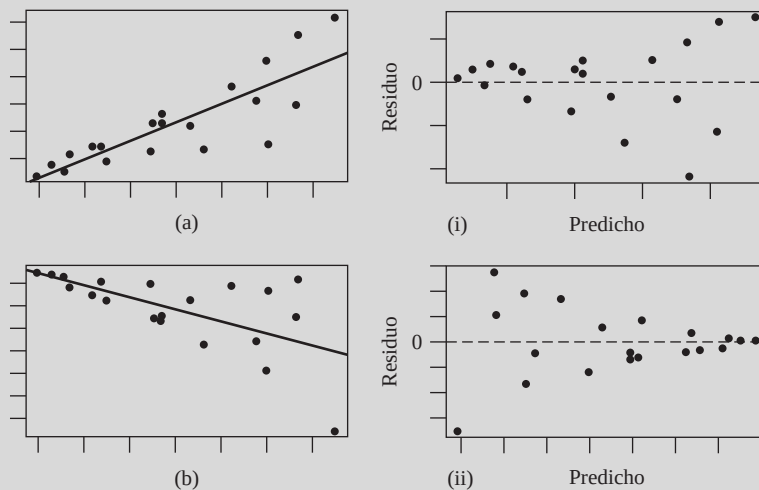
- ¿Consideraría el punto marcado como «×» un *outlier*? Explique su respuesta.
- ¿Consideraría el punto marcado como «×» un punto de apalancamiento? Explique su respuesta.
- Advirtiendo el pequeño cambio en las pendientes de la recta continua y discontinua, ¿consideraría el punto marcado como «×» una observación de influencia? Explique su respuesta.



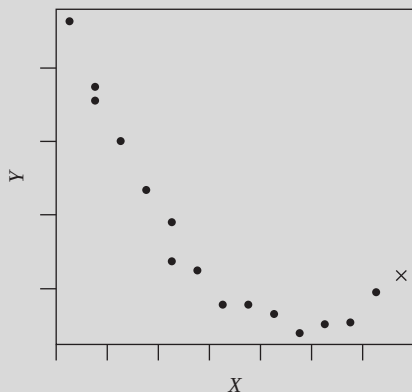
12.6.6 Las siguientes tres gráficas de residuos, (i), (ii) y (iii) se han generado tras ajustar rectas de regresión a los tres diagramas de dispersión siguientes: (a), (b) y (c). ¿Qué gráfica de residuos se corresponde con cada diagrama de dispersión? ¿Cómo lo sabe?



12.6.7 Las siguientes dos gráficas de residuos, (i) y (ii), se han generado tras ajustar rectas de regresión a los dos diagramas de dispersión siguientes, (a) y (b). ¿Qué gráfica de residuos se corresponde con cada diagrama de dispersión? ¿Cómo lo sabe?



12.6.8 Dibuje la gráfica de residuos que resultaría de ajustar una recta de regresión al siguiente diagrama de dispersión. Uno de los puntos está marcado con «x». Indique este punto en la gráfica de residuos.



12.6.9 (Ejercicio para computador) Unos investigadores midieron los diámetros de 20 árboles en un bosque húmedo del Amazonas central y utilizaron datación mediante ^{14}C para determinar sus edades. Los datos se presentan en la tabla siguiente²⁵. Considere el uso del diámetro, X , como predictor de la edad, Y .

- Realice un diagrama de dispersión de $Y = \text{edad}$ en función de $X = \text{diámetro}$ y ajuste una recta de regresión a los datos.
- Realice una gráfica de residuos de la regresión del apartado (a). Realice después una gráfica de probabilidad normal de los residuos. ¿Qué dicen esas gráficas sobre el uso de un modelo lineal y sobre los procedimientos de inferencia basados en la regresión?

Diámetro (cm)	Edad (años)	Diámetro (cm)	Edad (años)
180	1.372	115	512
120	1.167	140	512
100	895	180	455
225	842	112	352
140	722	100	352
142	657	118	249
139	582	82	249
150	562	130	227
110	562	97	227
150	552	110	172

- Tome el logaritmo de los valores de la edad. Realice un diagrama de dispersión de $Y = \log(\text{edad})$ en función de $X = \text{diámetro}$ y ajuste una recta de regresión a los datos.
- Represente un diagrama de residuos de la regresión del apartado (c). A continuación, realice una gráfica de probabilidad normal de los residuos. Basándose en estas gráficas, ¿parece apropiado un modelo de regresión utilizando la escala logarítmica del apartado (c)?

12.7 Precisión de la predicción (opcional)

En la Sección 12.4 aprendimos que un uso muy práctico de la regresión es la predicción. En esta sección distinguiremos entre la predicción del valor *medio* de Y para un valor de X particular y la predicción de un *solo* valor de Y para un valor de X particular. En concreto, compararemos las precisiones de esos dos tipos de predicción, que son muy diferentes.

Intervalos de confianza y de predicción

En el Ejemplo 12.4.6 utilizamos una recta de regresión para realizar una predicción: $\hat{y} = 197,17 - 2,51x$. Utilizando de nuevo esta recta podríamos predecir que la concentración *media* de arsénico en arroz a partir de plantas con concentración de silicio en la paja de 40 g/kg es $\hat{y} = 197,17 - 2,51(40) = 96,77 \mu\text{g/kg}$. ¿Qué sucedería si en vez de estimar la concentración media de arsénico de todas las plantas con esa concentración de silicio deseáramos predecir *la* concentración de arsénico de *una* planta particular cuya concentración de silicio en la paja fuera de 40 g/kg? Nuestra estimación sería la misma, $\hat{y} = 96,77 \mu\text{g/kg}$. Es decir, para estimar el valor medio de Y o un solo valor de Y para un valor particular de X , utilizamos la recta de regresión de la misma manera. Sin embargo, las precisiones de los estimadores son muy diferentes.

La predicción de un solo valor de Y es mucho menos precisa que la predicción de valor medio, debido a que, además de la incertidumbre en la recta de regresión (es decir, la incertidumbre de nuestros estimadores de la pendiente y la ordenada en el origen de la recta), existe también incertidumbre debida a la variabilidad inherente de los valores de Y que tienen el mismo valor de X . Por ejemplo, existe variabilidad entre las concentraciones de arsénico en arroz de todas las plantas cuya concentración de silicio en la paja es de 40 g/kg (de hecho, estimamos que esta variabilidad es s_e). Las

dos gráficas de la Figura 12.7.1 ilustran las diferencias de nuestras precisiones de predicción para los dos tipos de estimadores.

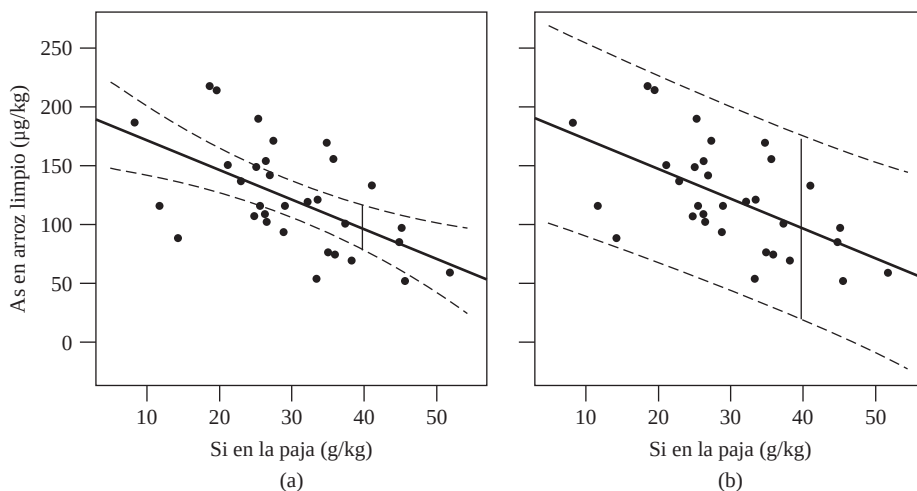


Figura 12.7.1 Bandas de confianza y de predicción del 95 % para las concentraciones de arsénico en arroz. La gráfica (a) muestra una banda de confianza del 95 % para las concentraciones medias predichas de arsénico cuando la concentración de silicio es de 40 g/kg. La gráfica (b) muestra una banda de confianza del 95 % para las concentraciones predichas de arsénico cuando la concentración de silicio es de 40 g/kg

La Figura 12.7.1(a) muestra una banda que representa todos los intervalos de confianza del 95 % para la predicción de niveles medios de arsénico, así como el intervalo específico para $X = 40$ g/kg marcado con una línea vertical. La banda de confianza refleja la incertidumbre asociada con la estimación de la pendiente y la ordenada en el origen de la recta de regresión. Nótese que los intervalos son más estrechos (más precisión) para concentraciones de silicio en la paja cerca del centro del conjunto de datos y mucho más amplios cerca de los valores de X de los extremos. Tenemos una confianza del 95 % en que la recta de regresión poblacional $\beta_0 + \beta_1 x$ esté dentro de esta banda. La mayor anchura de los intervalos cerca de los extremos es un reflejo de nuestra incertidumbre en la estimación de la pendiente de la recta de regresión. La anchura de la banda en el centro expresa nuestra incertidumbre de la altura global de la recta de regresión (con respecto a b_0).

Por el contrario, la Figura 12.7.1(b) muestra una banda que representa todos los intervalos del 95 % para la predicción de niveles de arsénico individuales. El intervalo de predicción específico para $X = 40$ se marca con una línea vertical. Nótese que esta banda es mucho más ancha en (b) que en (a). El Ejemplo 12.7.1 ilustra el uso de los intervalos de confianza y de predicción para la predicción en regresión.

Ejemplo 12.7.1 Arsénico en arroz

La Figura 12.7.1 muestra que para arroz con concentraciones de silicio en la paja de 40 g/kg, el intervalo de confianza del 95 % para la concentración media de arsénico es aproximadamente de 75 a 125 µg/kg. En otras palabras, tenemos una confianza del 95 % en que la concentración media de arsénico en arroz para plantas con concentración de silicio en la paja de 40 g/kg está entre 75 y 125 µg/kg. Por otra parte, utilizando intervalo de predicción estimamos que el 95 % de las plantas con concentraciones de silicio en la paja de 40 g/kg tendrán concentraciones de arsénico en arroz aproximadamente entre 25 y 175 µg/kg. ■

Recuérdese que la recta de regresión se puede interpretar como una «recta de promedios» y las medidas individuales necesariamente se separarán de este promedio. Estas gráficas nos muestran que tenemos mucha menos certeza al decir «el arroz de plantas con una cantidad X de silicio en la paja tendrá una cantidad Y de arsénico» que al decir «el arroz de plantas con una cantidad X de silicio en la paja tendrá, *en promedio*, una cantidad Y de arsénico».

Cálculo de los intervalos

Consideremos la predicción de $\mu_{Y|X=x^*}$ o de $Y|X=x^*$, es decir, la predicción de la media o del valor de Y real cuando $X=x^*$. El intervalo de confianza del 95 % para $\mu_{Y|X=x^*}$ está dado por

$$\hat{y} \pm t_{0,025} s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1)s_x^2}}$$

y el intervalo de confianza del 95 % para $Y|X=x^*$ está dado por

$$\hat{y} \pm t_{0,025} s_e \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1)s_x^2}}$$

donde el valor crítico $t_{0,025}$ se determina de la distribución t de Student con $gl = n - 2$.

Aunque estas dos fórmulas son muy similares, nótese el «1» extra bajo la raíz cuadrada en la fórmula del intervalo de predicción. Este «1» tiene en cuenta la variabilidad añadida asociada con intentar realizar una predicción de un valor individual en vez de una media poblacional.

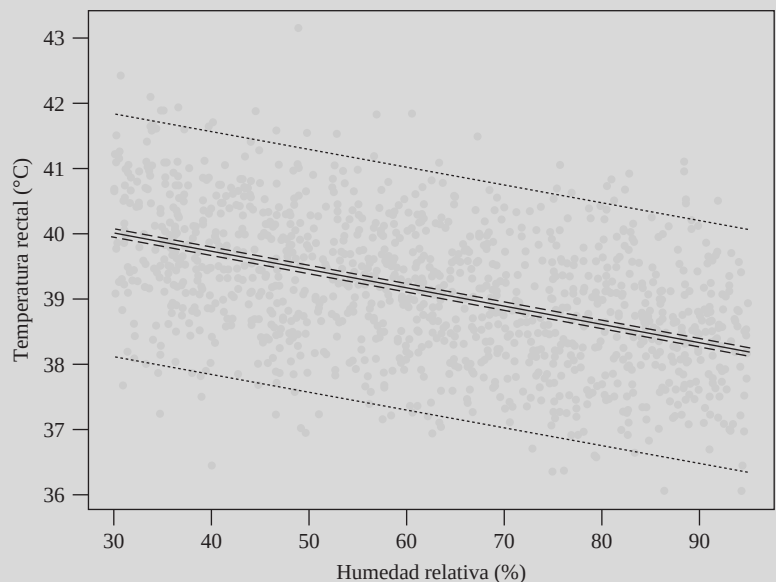
Como hemos visto en la Figura 12.7.1, tanto los intervalos de confianza como los de predicción son más amplios cuando realizamos predicciones lejos del centro de los datos. Ambas fórmulas tienen en cuenta esta incertidumbre adicional mediante el término $\frac{(x^* - \bar{x})^2}{(n-1)s_x^2}$. Este término será grande cuando x^* esté lejos de $\bar{x}$, aumentando la anchura del intervalo. Nótese que cuando $x^* = \bar{x}$ la fórmula del intervalo de confianza se reduce a una forma muy familiar: $\hat{y} \pm t_{0,025} \left(\frac{s_e}{\sqrt{n}} \right)$, que es muy similar a la fórmula del intervalo de confianza para la media poblacional presentada en el Capítulo 6.

La mayoría del *software* estadístico puede calcular y representar fácilmente las bandas de confianza y de predicción.

Ejercicios 12.7.1-12.7.3

12.7.1 En un estudio sobre el estrés de las vacas, unos investigadores midieron la temperatura rectal (°C) (Y) y la humedad relativa (X) de 1.280 vacas lactantes²⁶. La gráfica siguiente muestra los datos y la recta de regresión (línea continua). Hay otras dos parejas de líneas en esta gráfica: de rayas y de puntos. Una pareja de líneas muestra la banda de confianza del 95 % y la otra la banda de predicción del 95 %.

- ¿Qué pareja de líneas corresponde a la banda de confianza? ¿Qué nos dice esta banda?
- ¿Qué pareja de líneas corresponde a la banda de predicción? ¿Qué nos dice esta banda?
- Si el conjunto de datos fuera más pequeño, explique qué sucedería con esas bandas. ¿Serían bandas más amplias o más estrechas alrededor de la recta de regresión?

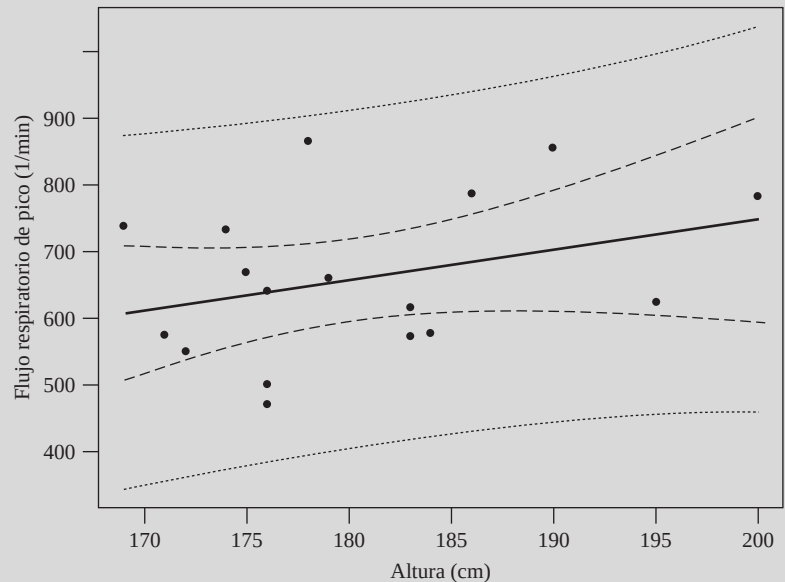


12.7.2 (Continuación del Ejercicio 12.7.1)

Suponga que se incluyen 5.000 vacas adicionales en la muestra y que se realiza una gráfica similar de los datos, recta de regresión y bandas de confianza y predicción con esta nueva muestra más grande. ¿Sería mucho más estrecha la banda de predicción? Explique su razonamiento.

12.7.3 La gráfica siguiente muestra la recta de regresión un las bandas de confianza y predicción del 95 % para los datos de flujo respiratorio de pico del Ejercicio 12.3.8.

- (a) Utilizando la gráfica para justificar su respuesta, ¿sería muy sorprendente encontrar un individuo con una altura de 195 cm y una velocidad de flujo de pico por encima de 900 l/min?
- (b) Utilizando la gráfica para justificar su respuesta, ¿sería sorprendente encontrar un grupo grande de individuos con una altura de 195 cm y que tuvieran velocidad de flujo de pico por encima de 900 l/min?



12.8 Perspectiva

Para poner en perspectiva los métodos del Capítulo 12, presentaremos su relación con métodos descritos en capítulos anteriores, y con métodos que se pueden incluir en un segundo curso de estadística. Empezaremos relacionando la regresión con los métodos de los Capítulos 7 y 11.

Regresión y test t

Cuando hay varios valores de Y para cada uno de dos valores de X , se podrían analizar los datos con un test t de dos muestras o con un análisis de regresión. Cada planteamiento utiliza los datos para estimar la media condicional de Y para cada X fijo. Estos parámetros se estiman mediante la recta ajustada $b_0 + b_1x$ en el planteamiento de regresión y mediante las medias muestrales individuales $\bar{Y}$ en el planteamiento del test t . Para contrastar la hipótesis nula de que no hay dependencia de Y con respecto a X , cada planteamiento expresa la hipótesis nula en sus propios términos. El ejemplo siguiente ilustra los planteamientos.

Ejemplo 12.8.1 Tolueno y cerebro

En el Capítulo 7 analizamos datos sobre concentraciones de norpinefrina (NE) en los cerebros de siete ratas expuestas a tolueno y en cinco ratas de control. Los datos se reproducen en la Tabla 12.8.1.

En el Capítulo 7 se contrastó la hipótesis nula

$$H_0: \mu_1 - \mu_2 = 0$$

Utilizando el test t de dos muestras (no conjunto). El estadístico de contraste fue

$$t_s = \frac{(540,83 - 444,20) - 0}{41,195} = 2,346$$

Tabla 12.8.1 Concentraciones de NE (ng/g)		
	Tolueno	Control
	543	535
	523	385
	431	502
	635	412
	564	387
	549	
n	6	5
$\bar{y}$	540,83	444,20
s	66,12	69,64

Estos datos se podrían analizar utilizando un test t conjunto (o, de forma equivalente, con análisis de varianza). La varianza conjunta es

$$s_{\text{conjunta}}^2 = \frac{(6 - 1)66,12^2 + (5 - 1)69,64^2}{(6 + 5 - 2)} = 4.584,24 = 67,71^2$$

y el ET conjunto es

$$ET_{\text{conjunto}} = 67,71 \sqrt{\frac{1}{6} + \frac{1}{5}} = 41,00$$

Esto conduce al estadístico de contraste

$$t_s = \frac{(540,83 - 444,20) - 0}{41,00} = 2,357$$

que no es muy diferente del resultado del test t no conjunto.

Estos datos se pueden analizar también con un modelo de regresión. Para utilizar regresión, definimos una **variable indicadora** (una variable que indica la pertenencia a un grupo) como sigue. Sea $X = 0$ para observaciones en el grupo de control y sea $X = 1$ para observaciones en el grupo del tolueno. Entonces, podemos representar gráficamente los datos con un diagrama de dispersión, como muestra la Figura 12.8.1.

Podemos analizar los datos del diagrama de dispersión con el modelo lineal

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

que establece que $\mu_{Y|X} = \beta_0 + \beta_1 X$.

El modelo lineal indica que para la ratas del grupo de control, la concentración de NE media (poblacional) está dada por

$$\mu_{Y|X=0} = \beta_0 + \beta_1(0) = \beta_0$$

Y para las ratas del grupo del tolueno, la concentración de NE está dada por

$$\mu_{Y|X=1} = \beta_0 + \beta_1(1) = \beta_0 + \beta_1$$

La diferencia entre las medias de los dos grupos es β_1 . Por tanto, la hipótesis nula

$$H_0: \mu_{Y|X=0} - \mu_{Y|X=1} = 0$$

es equivalente a la hipótesis nula

$$H_0: \beta_1 = 0$$

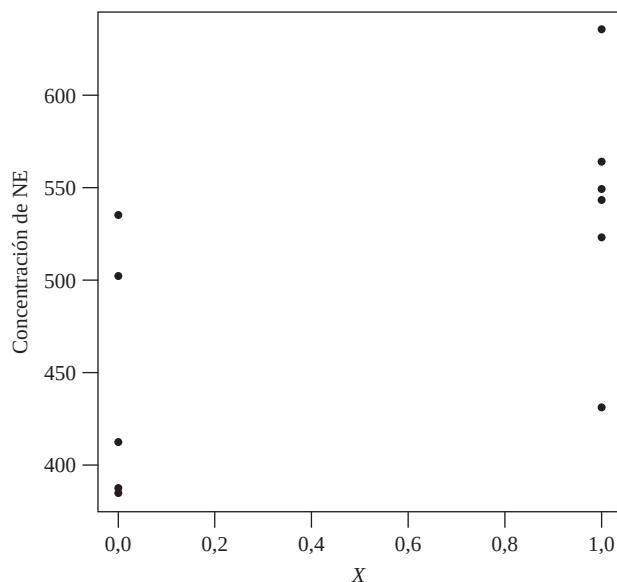


Figura 12.8.1 Datos de concentración de NE. $X = 0$ representa al grupo de control; $X = 1$ representa al grupo del tolueno

La recta de regresión ajustada es $\hat{y} = 444,2 + 96,63x$. Nótese que cuando $X = 0$, la recta de regresión ajustada da un valor de $\hat{y} = 444,2$, que es la media muestral del grupo de control. Cuando $X = 1$, la recta de regresión ajustada da un valor de $\hat{y} = 444,2 + 96,63 = 540,83$, que es la media muestral del grupo del tolueno. Es decir, el valor muestral de la pendiente es igual al cambio en las medias muestrales cuando se pasa de un grupo de control ($X = 0$) al grupo del tolueno ($X = 1$), como se muestra en la Figura 12.8.2.

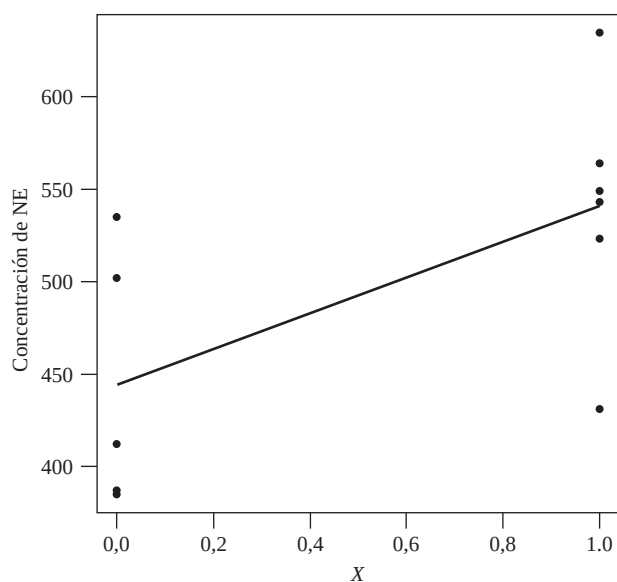


Figura 12.8.2 Datos de concentración de NE con la recta de regresión añadida

El estadístico de contraste para contrastar la hipótesis $H_0: \beta_1 = 0$ es

$$t_s = \frac{96,63}{41,0} = 2,36$$

Este estadístico es idéntico al estadístico del test de conjunto de dos muestras previo. (Nótese que el análisis de regresión asume que $\sigma_{Y|X} = \sigma_e$ es constante. Por tanto, la regresión es similar al test t conjunto, en vez de al test t no conjunto). La siguiente salida de computador muestra los coeficientes para la recta de regresión ajustada, así como el estadístico t .

The regression equations is
NE=444+96.6X

Predictor	Coef	SE Coef	T	P
Constant	444.20	30.28	14.67	0.000
X	96.63	41.00	2.36	0.043

S=67.7049 R-Sq=38.2% R-Sq(adj) = 31.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	25467	25467	5.56	0.043
Residual Error	9	41256	4584		
Total	10	66723			

El ejemplo siguiente compara el planteamiento de regresión y el planteamiento de dos muestras en un conjunto de datos en el que (a diferencia del Ejemplo 12.8.1) X varía dentro de las muestras y entre las muestras.

Ejemplo 12.8.2 Presión sanguínea y calcio en plaquetas

En el Ejemplo 12.5.3 presentamos medidas de presión sanguínea (X) y calcio en plaquetas (Y) de 38 sujetos. Realmente, el estudio incluía dos grupos de sujetos: 38 voluntarios con presión sanguínea normal, seleccionados entre el personal de laboratorio de un hospital y otras personas que no eran pacientes, y 45 pacientes con diagnóstico de presión sanguínea alta. La Tabla 12.8.2 resume las medidas de calcio en plaquetas en los dos grupos y la Figura 12.8.3 muestra las medidas de presión sanguínea y de calcio de los 83 sujetos⁴.

Las dos formas de analizar los datos son (1) como dos muestras independientes y (2) mediante un análisis de regresión. Para contrastar una relación entre la presión sanguínea y el calcio en plaquetas (1) se puede aplicar un test t de dos muestras con $H_0: \mu_1 = \mu_2$ a la Tabla 12.8.2; (2) se puede aplicar un test de regresión con $H_0: \beta_1 = 0$ a los datos de la Figura 12.8.3. El estadístico t de dos muestras (no conjunto) es $t_s = 11,2$ y el estadístico t de la regresión es $t_s = 20,8$. Ambos son altamente significativos, pero el último lo es más porque el análisis de regresión extrae más información de los datos.

Para estos datos, el planteamiento de regresión es más iluminador y convincente que el planteamiento de dos muestras. La Figura 12.8.3 sugiere que el calcio en plaquetas está correlacionado con la presión sanguínea, no solo entre, sino también dentro de los dos grupos. Los análisis de regresión relevantes deben incluir (1) un contraste de la correlación dentro de cada grupo separadamente (como en los Ejemplos 12.2.3 y 12.5.3); (2) un contraste de la correlación global (como en el párrafo anterior); (3) un contraste para ver si las rectas de regresión de los dos grupos son idénticas (utilizando métodos que no se describen en este libro).

Tabla 12.8.2 Calcio en plaquetas (nM) en dos grupos de sujetos		
	Presión sanguínea normal	Presión sanguínea alta
$\bar{y}$	107,9	168,2
s	16,1	31,7
n	38	45

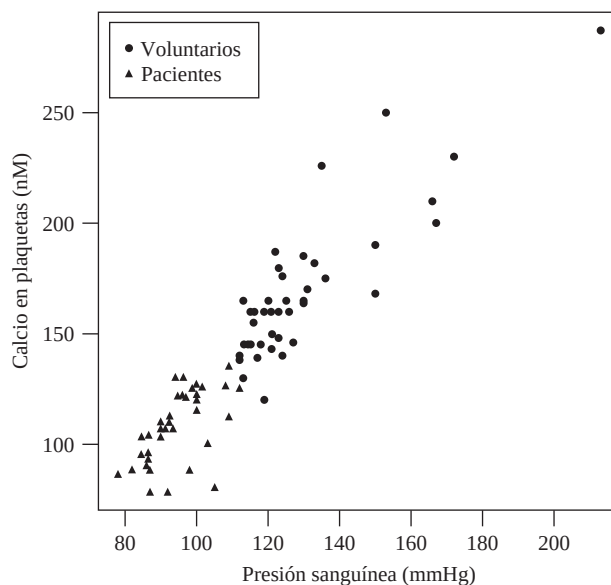


Figura 12.8.3 Presión sanguínea y calcio en plaquetas para 83 sujetos

Aparte de los contrastes formales, nótese la ventaja del diagrama de dispersión como herramienta para entender los datos y comunicar los resultados. La Figura 12.8.3 proporciona un testimonio elocuente de la realidad de la relación entre la presión sanguínea y el calcio en plaquetas. (Sin embargo, resaltamos de nuevo que una relación «real» no es necesariamente una relación causal. Es más, incluso si la relación fuera causal, los datos no indican la dirección de la causalidad, es decir, si el calcio alto produce presión sanguínea alta, o viceversa)*. ■

El Ejemplo 12.8.2 ilustra un principio general: si existe información cuantitativa sobre una variable X , en general es mejor utilizar esta información que ignorarla.

Extensiones de mínimos cuadrados

Hemos visto que el método clásico de ajuste de una recta a unos datos se basa en el criterio de mínimos cuadrados. Este criterio versátil se puede aplicar en muchos otros problemas estadísticos. Por ejemplo, en la **regresión curvilínea**, se utiliza el criterio de mínimos cuadrados para ajustar relaciones curvilíneas como

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

Otra aplicación es la **regresión y correlación múltiple**, en la que se usa el criterio de mínimos cuadrados para ajustar una ecuación que relaciona Y con varias variables X : X_1 , X_2 , etc. Por ejemplo,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

El ejemplo siguiente ilustra tanto la regresión múltiple como la curvilínea.

Ejemplo 12.8.3 Colesterol en el suero y presión sanguínea

Como parte de un gran estudio sobre la salud, se realizaron a 2.599 hombres varias medidas sobre presión sanguínea, química de la sangre y estado físico²⁷. Los investigadores encontraron una correlación positiva entre la presión sanguínea y el colesterol en el suero ($r = 0,23$ para la presión sanguínea sistólica). Pero la presión sanguínea y

* De hecho, los autores del estudio remarcan que «queda como posibilidad ... que el incremento en la concentración de calcio intracelular sea una consecuencia en vez de una causa de la presión sanguínea alta».

el colesterol en suero estaban también relacionados con la edad y el estado físico. Para esclarecer las relaciones, los investigadores utilizaron el método de mínimos cuadrados para ajustar la siguiente ecuación:

$$Y = b_0 + b_1X_1 + c_1X_1^2 + b_2X_2 + b_3X_3 + b_4X_4$$

siendo

Y = Presión sanguínea sistólica.

X_1 = Edad.

X_2 = Colesterol en suero.

X_3 = Glucosa en sangre.

X_4 = Índice ponderal (altura dividida por la raíz cúbica del peso).

Nótese que la regresión es curvilínea con respecto a la edad (X_1) y lineal en las otras variables X .

Aplicando análisis de regresión y correlación múltiple, los investigadores determinaron que había muy poca o ninguna correlación entre la presión sanguínea y el colesterol en suero, después de tener en cuenta cualquier relación entre la presión sanguínea y la edad e índice ponderal. Concluyeron que la correlación observada entre el colesterol en suero y la presión sanguínea es una consecuencia indirecta de la correlación de cada uno de ellos con la edad y el estado físico. ■

Regresión y correlación no paramétrica robusta

Hemos presentado los métodos clásicos de mínimos cuadrados para el análisis de regresión y correlación. Existen también muchos métodos modernos excelentes que no están basados en el criterio de mínimos cuadrados. Algunos de estos métodos son *robustos*, es decir, funcionan incluso si las distribuciones condicionales de Y dado X tienen colas largas y desordenadas o *outliers*. Los métodos no paramétricos asumen pocas cosas o ninguna sobre la forma de dependencia (lineal o curvilínea) de Y con respecto a X , o sobre la forma de las distribuciones condicionales.

Análisis de covarianza

Algunas veces las ideas de la regresión pueden aumentar mucho la potencia de un análisis de datos, incluso aunque la relación entre X e Y no sea el interés principal. Sigue a continuación un ejemplo.

Ejemplo 12.8.4 Tamaño de cabeza de orugas

¿Puede afectar la dieta al tamaño de la cabeza de las orugas? El efecto es creíble, debido a que los músculos de masticación de las orugas ocupan una gran parte de la cabeza. Para estudiar el efecto de la dieta, un biólogo alimentó a orugas (*Pseudaletia unipuncta*) con tres dietas diferentes: dieta 1, una dieta suave artificial; dieta 2, hierba suave, y dieta 3, hierba dura. Midió el tamaño de la cabeza y del cuerpo completo del estado final de desarrollo de la larva. Los resultados se muestran en la Figura 12.8.4, donde $Y = \ln(\text{peso de la cabeza})$ se representa con respecto a $X = \ln(\text{tamaño del cuerpo})$, con símbolos diferentes para las tres dietas²⁸. Nótese que el efecto de la dieta es impactante. Virtualmente no existe solapamiento entre los tres grupos de puntos. Pero si ignoráramos X y consideráramos solamente Y , como se muestra en la Figura 12.8.5, el efecto de la dieta sería mucho menos pronunciado. ■

El Ejemplo 12.8.4 muestra cómo la comparación de varios grupos con respecto a una variable Y se puede fortalecer utilizando información de una variable auxiliar X que esté correlacionada con Y . Un método clásico de análisis estadístico para estos datos es el **análisis de covarianza**, que se basa en ajustar rectas de regresión a los datos (X, Y) . Pero incluso sin esta técnica formal, los investigadores pueden a menudo clarificar la interpretación de los datos representando simplemente un diagrama de dispersión como el de la Figura 12.8.4. Dibujar los datos en función de X tiene el efecto visual de eliminar la parte de la variabilidad de Y debida a X , causando que el efecto del tratamiento resalte más claramente contra la variación del fondo residual.

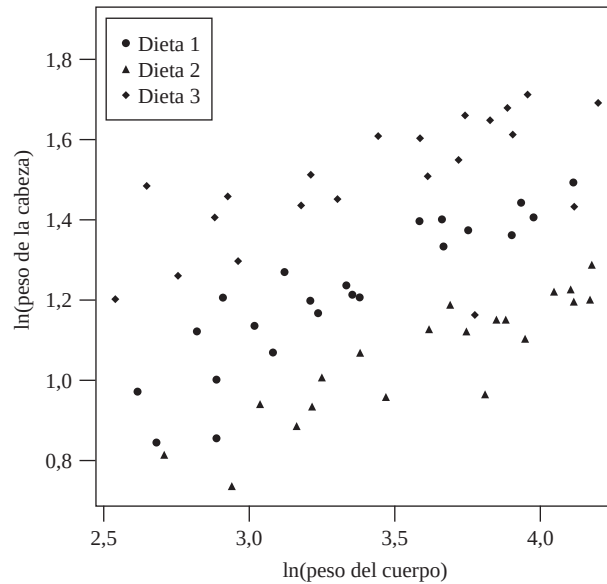


Figura 12.8.4 Peso de la cabeza en función de peso del cuerpo (en escala logarítmica) para orugas con tres dietas diferentes

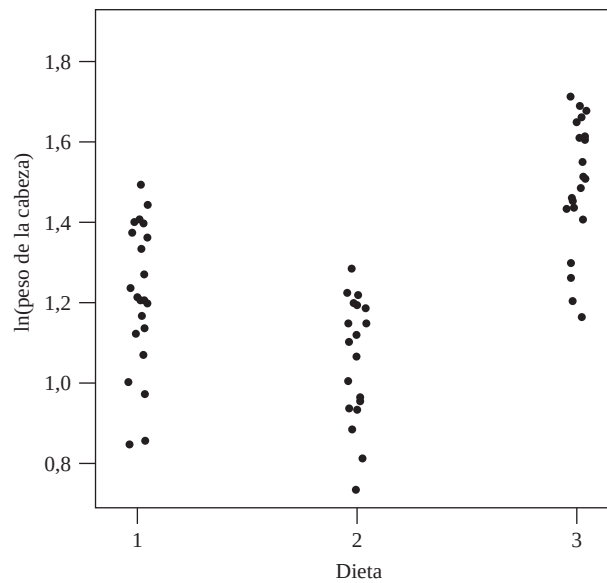


Figura 12.8.5 Peso de la cabeza (en escala logarítmica) para orugas con tres dietas diferentes

Regresión logística

La regresión y la correlación se utilizan para analizar la relación entre dos variables cuantitativas, X e Y . Algunas veces aparecen datos en los que una variable cuantitativa X se utiliza para predecir la respuesta de una variable categórica Y . Por ejemplo, podríamos desear usar X = nivel de colesterol como predictor de si una persona tendrá o no un ataque cardíaco. En este caso podemos definir una variable Y que valiera 1 si una persona tiene un ataque cardíaco y 0 si no lo

tiene. Podríamos entonces estudiar cómo depende Y de X . Cuando la variable de respuesta es dicotómica como en este caso, se puede utilizar una técnica conocida como **regresión logística** para modelar la relación. Por ejemplo, se podría utilizar regresión logística para modelar cómo depende de la presión sanguínea la probabilidad de tener un ataque cardíaco.

El Ejemplo 12.8.5 proporciona una visión más detallada del uso de la regresión logística.

Ejemplo 12.8.5 Cáncer de esófago

El cáncer de esófago es una enfermedad seria y muy agresiva. Unos científicos realizaron un estudio de 31 pacientes con cáncer de esófago en el que estudiaron la relación entre el tamaño del tumor que tenía el paciente y si el cáncer se había extendido (metástasis) o no a los ganglios linfáticos del paciente. En este estudio la variable de respuesta es dicotómica: $Y = 1$ si el cáncer se había extendido a los ganglios linfáticos e $Y = 0$ si no se había extendido. La variable de predicción es el tamaño (la máxima dimensión, en cm) del tumor encontrado en el esófago. Los datos se recogen en la Tabla 12.8.3 y se presentan en forma gráfica en la Figura 12.8.6²⁹.

Tabla 12.8.3 Datos de cáncer de esófago					
Número de paciente	Tamaño del tumor (cm), X	Metástasis en los ganglios linfáticos, Y	Número de paciente	Tamaño del tumor (cm), X	Metástasis en los ganglios linfáticos, Y
1	6,5	1	17	6,2	1
2	6,3	0	18	2,0	0
3	3,8	1	19	9,0	1
4	7,5	1	20	4,0	0
5	4,5	1	21	3,0	1
6	3,5	1	22	6,0	1
7	4,0	0	23	4,0	0
8	3,7	0	24	4,0	0
9	6,3	1	25	4,0	0
10	4,2	1	26	5,0	1
11	8,0	0	27	9,0	1
12	5,2	1	28	4,5	1
13	5,0	1	29	3,0	0
14	2,5	0	30	3,0	1
15	7,0	1	31	1,7	0
16	5,3	0			

La idea de la regresión logística es modelar la relación entre X e Y ajustando una curva de respuesta que esté siempre entre 0 y 1. Acotando los valores entre 0 y 1, el modelo de regresión logística se puede utilizar para estimar la probabilidad de que $Y = 1$ (es decir, metástasis) para un valor dado de X (es decir, tamaño del tumor). Por tanto, a diferencia de la regresión lineal, en la que modelamos Y como una función lineal de X (cuyos valores no están entre 0 y 1), con la regresión logística modelamos la relación entre X e Y mediante una curva en forma de «S», como se muestra en la Figura 12.8.7.

Una forma de empezar a interpretar los datos es formar grupos basados en su tamaño, X , y calcular para cada grupo la proporción de los valores de Y que son 1. (Esto es en cierta manera análogo a obtener la gráfica de promedios descrita en la Sección 12.3, excepto porque aquí agrupamos los puntos de datos con diferentes valores

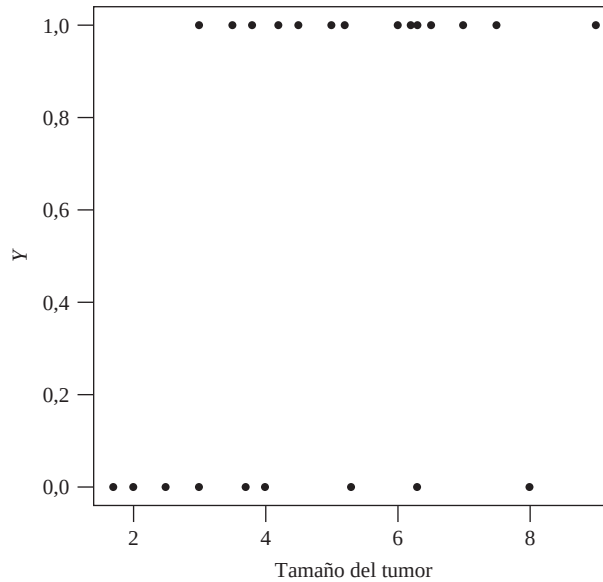


Figura 12.8.6 Metástasis en los ganglios linfáticos, Y , en función del tamaño del tumor, X

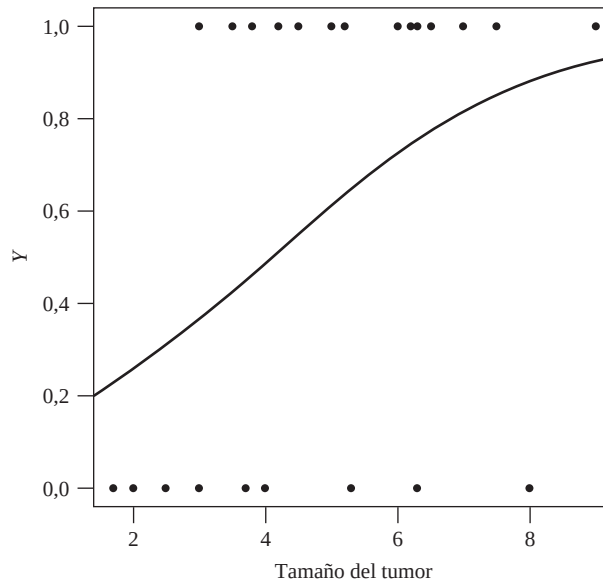


Figura 12.8.7 Metástasis en los ganglios linfáticos, Y , en función del tamaño del tumor, X , con una curva suave añadida

de X). La Tabla 12.8.4 proporciona un resumen de este tipo, que se muestra gráficamente en la Figura 12.8.8. Nótese que la proporción de unos (es decir, la proporción de pacientes en los que el cáncer ha causado metástasis) aumenta a medida que el tamaño del tumor aumenta (excepto para la última categoría de (7,5, 9], que solo tiene tres casos).

Podemos ajustar una función continua y suave a los datos, para suavizar los porcentajes de la última columna de la Tabla 12.8.4. Podemos imponer también la condición de que la función sea monótona creciente, lo que significa que la probabilidad de metástasis ($Y = 1$) crece estrictamente cuando el tamaño del tumor aumenta. Para ello,

Tabla 12.8.4 Datos de cáncer de esófago en grupos				
Intervalo de tamaños	Puntos con $Y = 1$	Puntos con $Y = 0$	Fracción $Y = 1$	Proporción $Y = 1$
(1,5, 3,0]	2	4	2/6	0,33
(3,0, 4,5]	5	6	5/11	0,45
(4,5, 6,0]	4	1	4/5	0,80
(6,0, 7,5]	5	1	5/6	0,83
(7,5, 9,0]	2	1	2/3	0,67

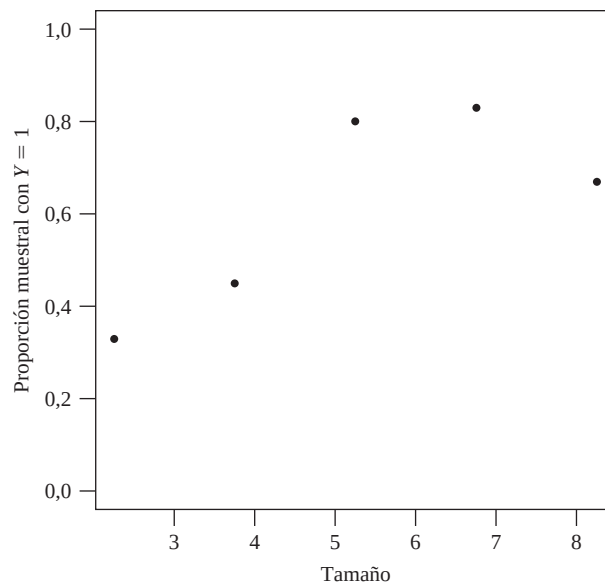


Figura 12.8.8 Proporción muestral de pacientes con metástasis en los ganglios linfáticos ($Y = 1$), agrupados por tamaño del tumor, X

usamos un computador para ajustar una **función de respuesta logística***. La función de respuesta logística ajustada para los datos sobre cáncer de esófago es

$$\Pr\{Y = 1\} = \frac{e^{-2,086 + 0,5117 \times \text{tamaño}}}{1 + e^{-2,086 + 0,5117 \times \text{tamaño}}}$$

Por ejemplo, supongamos que el tamaño de un tumor es de 4,0 cm. Entonces la probabilidad predicha de que el cáncer haya desarrollado metástasis es

$$\frac{e^{-2,086 + 0,5117(4)}}{1 + e^{-2,086 + 0,5117(4)}} = \frac{e^{-0,0392}}{1 + e^{-0,0392}} = \frac{0,96156}{1 + 0,96156} = 0,49$$

* El ajuste de un modelo logístico es bastante más complicado que el ajuste de un modelo lineal. Se utiliza comúnmente una técnica denominada estimación por máxima verosimilitud, con la ayuda de un ordenador.

Por otra parte, supongamos que el tamaño del tumor es de 8,0 cm. Entonces la probabilidad predicha de que el cáncer haya desarrollado metástasis es

$$\frac{e^{-2,086+0,5117 \times 8}}{1 + e^{-2,086+0,5117 \times 8}} = \frac{e^{2,0076}}{1 + e^{2,0076}} = \frac{7,4454}{1 + 7,4454} = 0,88$$

Podemos calcular una probabilidad predicha de que $Y = 1$ para cada valor de X . La Figura 12.8.9 muestra una gráfica de estas predicciones que tienen, hablando en términos generales, forma de S. ■

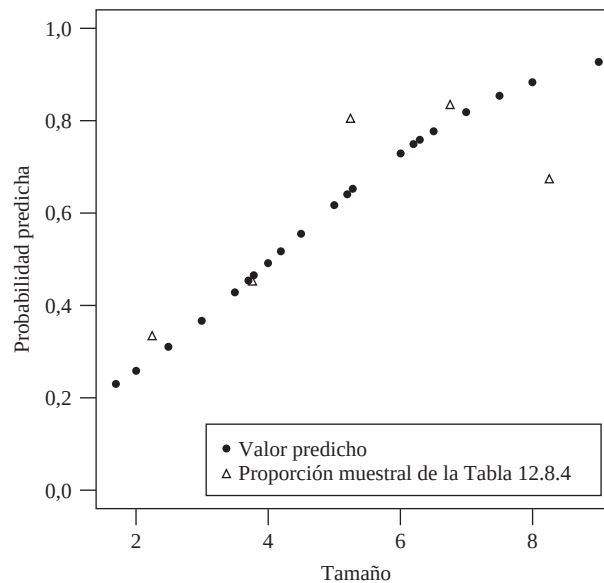


Figura 12.8.9 Proporción predicha de que $Y = 1$ en función del tamaño del tumor, X , con las proporciones muestrales de la Tabla 12.8.4

La forma de S de la curva logística es fácil de ver si ampliamos el intervalo de valores de X , como se muestra en la Figura 12.8.10. Cuando X crece, la curva logística se aproxima, sin llegar nunca, a 1. Asimismo, si extendiéramos la curva en la región donde X es menor que 0 veríamos que a medida que X se hace más y más pequeño, la curva logística se aproxima, sin llegar nunca, a 0. (Por supuesto, en el escenario del Ejemplo 12.8.5 no tiene sentido hablar de tamaños de tumor que sean negativos. Por tanto, solo representamos la curva logística para valores positivos de X).

En general, si tenemos una función de respuesta logística

$$\Pr\{Y = 1\} = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

con b_1 positivo, entonces cuando X crece, $\Pr\{Y = 1\}$ se aproxima a 1 y cuando X disminuye, $\Pr\{Y = 1\}$ se aproxima a 0. Por tanto, a diferencia de un modelo de regresión lineal, una curva logística tiene sus valores entre 0 y 1, lo que la hace apropiada para modelar una probabilidad de respuesta.

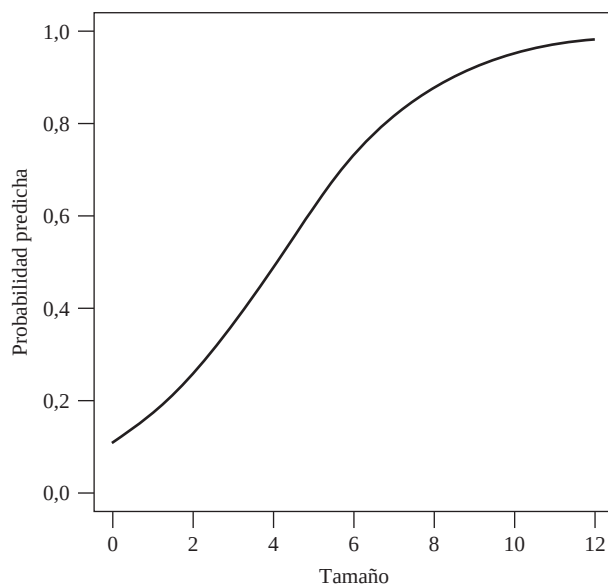


Figura 12.8.10 Función de respuesta logística para los datos de cáncer, representada en un intervalo más amplio

12.9 Resumen de fórmulas

Para disponer de una referencia conveniente, resumimos las fórmulas presentadas en el Capítulo 12.

Coeficiente de correlación

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x - \bar{x}}{s_x} \right) \left(\frac{y - \bar{y}}{s_y} \right)$$

Propiedad 12.3.1:

$$r^2 \approx \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

Recta de regresión ajustada

siendo

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = r \times \left(\frac{s_y}{s_x} \right)$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Residuos:

$$y_i - \hat{y}_i \quad \text{siendo} \quad \hat{y}_i = b_0 + b_1 x_i$$

Suma de cuadrados de residuos:

$$SC(\text{resid}) = \sum (y_i - \hat{y}_i)^2$$

Desviación típica residual:

$$s_e = \sqrt{\frac{SC(\text{resid})}{n-2}}$$

Inferencia

Error típico de b_1 :

$$ET_{b_1} = \frac{s_e}{s_x \sqrt{n-1}}$$

Intervalo de confianza del 95 % para β_1 :

$$b_1 \pm t_{0,025} ET_{b_1}$$

Contraste de $H_0: \beta_1 = 0$ o $H_0: \rho = 0$:

$$t_s = \frac{b_1}{ET_{b_1}} = r \sqrt{\frac{n-1}{1-r^2}}$$

Los valores críticos para el test y el intervalo de confianza se determinan a partir de la distribución t de Student con $gl = n - 2$.

Predicción

El intervalo de confianza para $\mu_{Y|X=x^*}$ está dado por

$$\hat{y} \pm t_{0,025} s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1)s_x^2}}$$

Un intervalo de confianza del 95 % de la predicción para $Y|X = x^*$ está dado por

$$\hat{y} \pm t_{0,025} s_e \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1)s_x^2}}$$

Los valores críticos de los intervalos se determinan a partir de la distribución t de Student con $gl = n - 2$.

Ejercicios 12.S.1-12.S.23

12.S.1 En un estudio sobre el grillo Mormon (*Anabrus simplex*) se encontró que la correlación entre el peso del cuerpo de las hembras y el peso de los ovarios era de $r = 0,836$. La desviación típica de los pesos de los ovarios fue de 0,429 g. Suponiendo que es aplicable el modelo lineal, estime la desviación típica de los pesos de los ovarios de los grillos cuyo peso corporal es de 4 g³⁰.

12.S.2 En un estudio sobre pérdidas de cosechas debidas a la contaminación de aire, se cultivaron parcelas de judías Blue Lake en cámaras de techo abierto, y se fumigaron con diversas dosis de dióxido de azufre. Después de un mes de fumigación, se recogieron las plantas y se contabilizó el número total de vainas de judía en cada cámara. Los resultados se muestran en la tabla³¹.

Los cálculos preliminares produjeron los siguientes resultados.

$$\begin{aligned} \bar{x} &= 0,12 & \bar{y} &= 1,117 \\ s_x &= 0,11724 & s_y &= 0,31175 \\ r &= -0,8506 & SC(\text{resid}) &= 0,2955 \end{aligned}$$

	X = concentración de dióxido de azufre (ppm)			
	0	0,06	0,12	0,30
Y = cosecha (kg)	1,15	1,19	1,21	0,65
	1,30	1,64	1,00	0,76
	1,57	1,13	1,11	0,69
Media	1,34	1,32	1,11	0,70

- Calcule la regresión lineal de Y respecto a X .
- Represente los datos y dibuje sobre la gráfica la recta de regresión.
- Calcule s_e . ¿Cuáles son las unidades de s_e ?

12.S.3 En referencia al Ejercicio 12.S.2:

- Suponiendo que el modelo lineal es aplicable, calcule estimaciones de la media y la desviación típica de una cosecha de judías expuesta a 0,24 ppm de dióxido de azufre.

(b) ¿Qué condición del modelo lineal parece dudosa para los datos de las judías?

12.S.4 En referencia al Ejercicio 12.S.2, considere la hipótesis nula de que la concentración del dióxido de azufre no tiene efecto en la cosecha. Suponiendo que se mantiene el modelo lineal, formule eso como una hipótesis sobre la verdadera recta de regresión. Utilice los datos para contrastar la hipótesis contra una alternativa direccional. Use $\alpha = 0,05$.

12.S.5 Otra forma de analizar los datos del Ejercicio 12.S.2 es tomar la media de cada tratamiento como la observación Y . Entonces los datos se resumían como indica la tabla siguiente.

	Dióxido de azufre X (ppm)	Cosecha media Y (kg)
	0,00	1,34
	0,06	1,32
	0,12	1,11
	0,30	0,70
Media	0,1200	1,1175
DT	0,12961	0,29714
	$r = -0,98666$ $SC(\text{resid}) = 0,007018$	

(a) Para la regresión de la cosecha media con respecto a X , calcule la recta de regresión y la desviación típica residual, y compárelas con los resultados del Ejercicio 12.S.2. Explique por qué la discrepancia no es sorprendente.

(b) ¿Qué proporción de la variabilidad de la cosecha media se explica por la relación lineal entre la cosecha media y el dióxido de azufre? Utilizando los datos del Ejercicio 12.S.5, ¿qué proporción de la variabilidad en la cosecha de una cámara individual se explica por la relación lineal entre la cosecha de una cámara individual y el dióxido de azufre? Explique por qué la discrepancia no es sorprendente.

12.S.6 En un estudio sobre el carbonero de cresta negra (*Parus bicolor*), un ecólogo capturó siete pájaros macho, midió la longitud de sus alas y otras características, después los marcó y los dejó en libertad. Durante el invierno siguiente, observó repetidamente a los pájaros marcados mientras buscaban insectos y semillas en las ramas de los árboles. En cada ocasión anotó el diámetro de la rama y calculó (con 50 observaciones) el diámetro de rama promedio para cada pájaro. Los resultados se muestran en la tabla siguiente³².

(a) Calcule s_e y especifique sus unidades. Verifique la relación aproximada entre s_Y y s_e , y r .

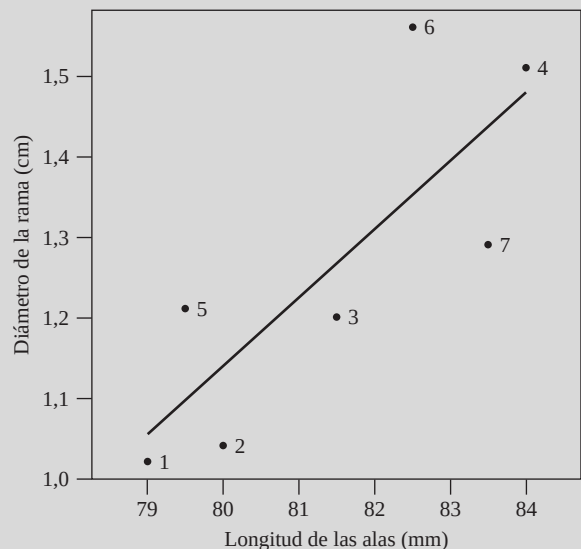
(b) ¿Proporcionan los datos evidencia suficiente para concluir que el diámetro de las ramas elegidas por los pája-

Pájaro	Longitud de las alas X (mm)	Diámetro de la rama Y (cm)
1	79,0	1,02
2	80,0	1,04
3	81,5	1,20
4	84,0	1,51
5	79,5	1,21
6	82,5	1,56
7	83,5	1,29
Media	81,429	1,2614
DT	1,98806	0,21035
	$r = 0,80335$ $SC(\text{resid}) = 0,09415$	

ros para buscar comida están correlacionados con la longitud de sus alas? Contraste una hipótesis adecuada contra una alternativa no direccional. Use $\alpha = 0,05$.

(c) El test del apartado (b) se basa en 7 observaciones, pero cada valor de diámetro de rama fue la media de 50 observaciones. Si contrastáramos la hipótesis del apartado (a) con los datos iniciales sin promediar, tendríamos 350 observaciones en lugar de solo 7. ¿Por qué no sería válido este planteamiento?

12.S.7 (Continuación del Ejercicio 12.S.6) Se presenta a continuación un diagrama de dispersión y una recta de regresión ajustada de los datos del Ejercicio 12.S.6. Los pájaros están etiquetados en la figura.



(a) ¿Qué pájaro/punto tiene el máximo residuo de regresión?

- (b) ¿Qué pájaro(s)/punto(s) tienen apalancamiento máximo?
- (c) ¿Hay muchos pájaros/puntos que sean de influencia?
- (d) Invente su propia observación de un pájaro de x = longitud de las alas e y = diámetro de la rama que sea un ejemplo de un *outlier* de la regresión.
- (e) Invente su propia observación de un pájaro de x = longitud de las alas e y = diámetro de la rama que sea un ejemplo de un punto de apalancamiento.

12.S.8 El Ejercicio 12.3.7 trata con datos sobre la relación entre la longitud corporal y la distancia de salto de ranas toro. La siguiente tabla muestra estos datos¹⁶.

Rana toro	Longitud X (mm)	Masa Y (g)
1	155	404
2	127	240
3	136	296
4	135	303
5	158	422
6	145	308
7	136	252
8	172	533,8
9	158	470
10	162	522,9
11	162	356
Media	149,636	373,427
DT	14,4725	104,2922

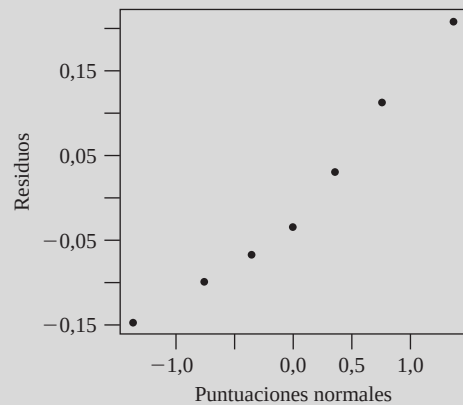
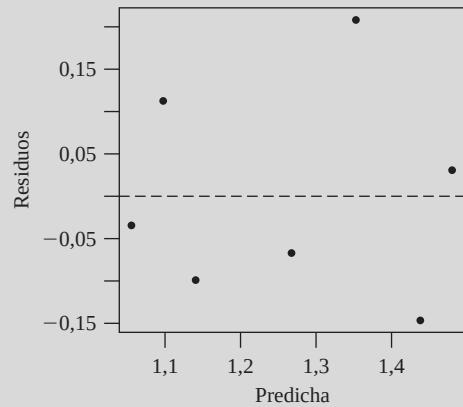
Los cálculos preliminares producen los siguientes resultados:

$$r = 0,90521 \quad SC(\text{resid}) = 19.642$$

- (a) Calcule la regresión lineal de Y con respecto a X .
- (b) Interprete el valor de la pendiente de la recta de regresión, b_1 , en el contexto de este ejercicio.
- (c) Calcule e interprete el valor de s_e en el contexto de este ejercicio.
- (d) Calcule e interprete el valor r^2 en el contexto de este ejercicio.

12.S.9 (Continuación del Ejercicio 12.S.8) Se presentan a continuación una gráfica de residuos y una gráfica de probabilidad de la regresión lineal de Y con respecto a X basada en los datos de masa de ranas toro del Ejercicio 12.S.8.

Utilice esas gráficas para comentar las condiciones requeridas para realizar una inferencia sobre la regresión. ¿Hay alguna razón para dudar sustancialmente del cumplimiento de esas condiciones?

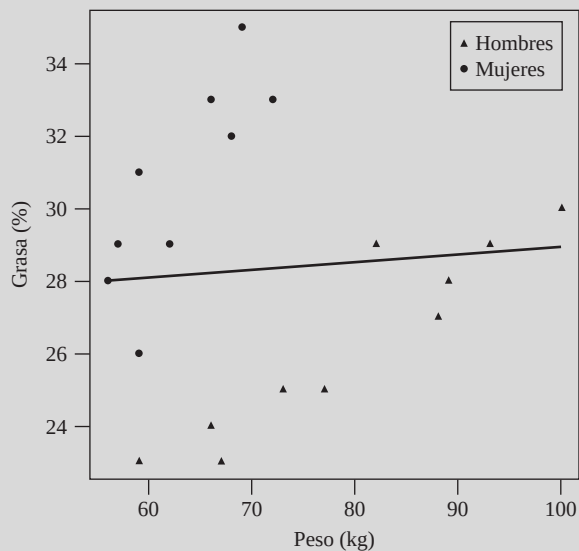


12.S.10 Un fisiólogo del ejercicio utilizó medidas de los pliegues cutáneos para estimar la grasa corporal total, expresada como porcentaje del peso corporal, para 19 participantes en un programa de ejercicio físico. Los porcentajes de grasa corporal y los pesos se muestran en la tabla siguiente³³.

Participante	Peso X (kg)	Grasa Y (%)	Participante	Peso X (kg)	Grasa Y (%)
1	89	28	11	57	29
2	88	27	12	68	32
3	66	24	13	69	35
4	59	23	14	59	31
5	93	29	15	62	29
6	73	25	16	59	26
7	82	29	17	56	28
8	77	25	18	66	33
9	100	30	19	72	33
10	67	23			

En la realidad, los participantes 1 a 10 eran hombres y los participantes 11 a 19 eran mujeres. Se presenta a continuación un resumen y una representación de los datos para hombres, mujeres y ambos sexos combinados.

Hombres ($n=10$)	Mujeres ($n=9$)	Ambos sexos ($n=19$)
$\bar{x} = 79,40$	$\bar{x} = 63,1$	$\bar{x} = 71,68$
$\bar{y} = 26,30$	$\bar{y} = 30,67$	$\bar{y} = 28,37$
$s_x = 13,2430$	$s_x = 5,7975$	$s_x = 13,1320$
$s_y = 2,6269$	$s_y = 2,8723$	$s_y = 3,4835$
$r = 0,9352$	$r = 0,8132$	$r = 0,0780$

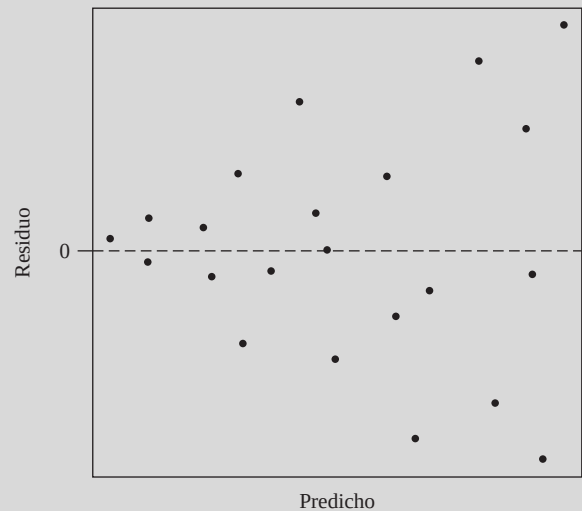


- Calcule separadamente las ecuaciones de la regresión de los hombres y de las mujeres.
- La ecuación de la recta de regresión ajustada para ambos sexos combinados es $\hat{y} = 26,88 + 0,021x$. ¿Cómo se compara la pendiente de esta recta con las pendientes calculadas en el apartado (a)? ¿Puede explicar la discrepancia?
- Examine los coeficientes de correlación para (i) los hombres, (ii) las mujeres y (iii) ambos sexos combinados. ¿Son coherentes esos valores con su razonamiento del apartado (b)?

12.S.11 En referencia a los datos de tasa de respiración del Ejercicio 12.3.6, construya un intervalo de confianza del 95 % para β_1 .

12.S.12 La gráfica siguiente representa los residuos del ajuste de un modelo de regresión a unos datos. Realice un dibujo del diagrama de dispersión de los datos que condujo a este diagrama de residuos. (Nota: hay dos diagramas de dis-

persión posibles, uno en el que b_1 es positivo y otro en el que b_1 es negativo.



12.S.13 Unos biólogos estudiaron la relación entre la frecuencia cardíaca de los embriones y la masa del huevo en 20 especies de pájaros. Obtuvieron que la frecuencia cardíaca, Y , tiene una relación lineal con el logaritmo de la masa del huevo, X . Los datos se muestran en la siguiente tabla³⁴.

Especies	Masa del huevo (g)	Log-(masa del huevo) X	Frecuencia cardíaca Y (latidos/min)
Diamante mandarín	0,96	-0,018	335
Diamante bengalí	1,10	0,041	404
Herrerillo de los pantanos	1,39	0,143	363
Golondrina de banco	1,42	0,152	298
Carbonero común	1,59	0,201	348
Carbonero variado	1,69	0,228	356
Gorrión molinero	2,09	0,320	335
Periquito	2,19	0,340	314
Avión común	2,25	0,352	357
Turco japonés	2,56	0,408	370
Estorninos de mejillas rojas	4,14	0,617	358
Cockatiel	5,08	0,706	300
Bulbul de orejas marrones	6,40	0,806	333
Paloma doméstica	17,10	1,233	247

(Continúa)

Especies	Masa del huevo (g)	Log-(masa del huevo) X	Frecuencia cardiaca Y (latidos/min)
Paloma de cola de milano	19,70	1,294	267
Paloma mensajera	19,80	1,297	230
Lechuza	20,10	1,303	219
Cuervo	20,50	1,312	297
Garza bueyera	27,50	1,439	251
Halcón borní	41,20	1,615	242
Media	9,94	0,690	311

Para estos datos la ecuación de regresión ajustada es

$$\hat{y} = 368,06 - 82,452x$$

y

$$SC(\text{resid}) = 15.748,6$$

- Interprete el valor de la ordenada en el origen de la recta de regresión, b_0 , en el contexto de este ejercicio.
- Interprete el valor de la pendiente de la recta de regresión, b_1 , en el contexto de este ejercicio.
- Calcule s_e y especifique sus unidades.
- Interprete el valor de s_e en el contexto de este ejercicio.

12.S.14 (Ejercicio para computador) La tabla que acompaña a este ejercicio presenta dos conjuntos de datos (A) y (B). Los valores de X son los mismos en ambos conjuntos de datos y se dan solo una vez.

	(A)	(B)		(A)	(B)
X	Y	Y	X	Y	Y
0,61	0,88	0,96	2,56	1,97	1,20
0,93	1,02	0,97	2,74	2,02	3,59
1,02	1,12	0,07	3,04	2,26	3,09
1,27	1,10	2,54	3,13	2,27	1,55
1,47	1,44	1,41	3,45	2,43	0,71
1,71	1,45	0,84	3,48	2,57	3,05
1,91	1,41	0,32	3,79	2,53	2,54
2,00	1,59	1,46	3,96	2,73	3,33
2,27	1,58	2,29	4,12	2,92	2,38
2,33	1,66	2,51	4,21	2,96	3,08

- Genere diagramas de dispersión de los dos conjuntos de datos.
- Para cada conjunto de datos (i) estime r visualmente y (ii) calcule r .

- Para el conjunto de datos (a), multiplique los valores de X por 10 y multiplique los valores de Y por 3 y sume 5. Vuelva a calcular r y compare el resultado con el valor antes de la transformación. ¿Cómo queda afectado r por la transformación lineal?
- Obtenga las ecuaciones de las rectas de regresión y verifique que dichas rectas son virtualmente idénticas para los dos conjuntos de datos (incluso aunque los coeficientes de correlación sean muy diferentes).
- Dibuje la recta de regresión sobre cada diagrama de dispersión.
- Construya un diagrama de dispersión en el que los dos conjuntos de datos estén superpuestos, utilizando diferentes símbolos en cada conjunto de datos.

12.S.15 (Ejercicio para computador) Este ejercicio muestra la potencia de los diagramas de dispersión para revelar características de los datos que pueden no resultar aparentes a partir de los cálculos ordinarios de la regresión lineal. La tabla que acompaña a este ejercicio muestra tres conjuntos de datos ficticios, A, B y C. Los valores de X son los mismos para cada conjunto de datos, pero los valores de Y son diferentes³⁵.

Conjunto de datos:	A	B	C
X	Y	Y	Y
10	8,04	9,14	7,46
8	6,95	8,14	6,77
13	7,58	8,74	12,74
9	8,81	8,77	7,11
11	8,33	9,26	7,81
14	9,96	8,10	8,84
6	7,24	6,13	6,08
4	4,26	3,10	5,39
12	10,84	9,13	8,15
7	4,82	7,26	6,42
5	5,68	4,74	5,73

- Verifique que la recta de regresión ajustada es casi la misma para los tres conjuntos de datos. ¿Son las desviaciones típicas residuales las mismas? ¿Son los valores de r los mismos?
- Construya un diagrama de dispersión para cada uno de los tres conjuntos de datos. ¿Qué indica cada gráfica sobre la idoneidad de la regresión lineal para cada conjunto de datos?
- Dibuje la recta de regresión ajustada sobre cada uno de los diagramas de dispersión.

12.S.16 (Ejercicio para computador) En un estudio farmacológico se asignaron aleatoriamente 12 ratas para recibir

una inyección de anfetamina con dos posibles niveles de dosificación o una inyección de suero salino. En la tabla que sigue se muestra el consumo de agua de cada animal (ml de agua por kg de peso corporal) durante las 24 horas siguientes a la inyección³⁶.

Dosis de anfetamina (ml/kg)		
0	1,25	2,5
122,9	118,4	134,5
162,1	124,4	65,1
184,1	169,4	99,6
154,9	105,3	89,0

- Calcule la recta de regresión del consumo de agua en función de la dosis de anfetamina, y calcule la desviación típica residual.
- Dibuje un diagrama de dispersión del consumo de agua en función de la dosis.
- Dibuje la recta de regresión sobre el diagrama de dispersión.
- Utilice la regresión lineal para contrastar la hipótesis de que la anfetamina no tiene efecto en el consumo de agua frente la alternativa de que la anfetamina tiende a reducir el consumo de agua. (Use $\alpha = 0,05$).
- Utilice análisis de varianza para contrastar la hipótesis de que la anfetamina no tiene efecto en el consumo de agua. (Use $\alpha = 0,05$). Compare el resultado con el resultado del apartado (d).
- ¿Qué condiciones son necesarias para la validez del test del apartado (d) pero no para la validez del test en el apartado (e)?
- Calcule la desviación típica conjunta del ANOVA, y compárela con la desviación típica residual calculada en el apartado (a).

12.S.17 (Ejercicio para computador) Consideremos los datos de árboles del Amazonas del Ejercicio 12.6.9. Los investigadores de este estudio estaban interesados en cómo la edad, Y , está relacionada con $X =$ «velocidad de crecimiento», donde la velocidad de crecimiento se define como diámetro/edad (es decir, cm de crecimiento al año).

- Cree la variable «velocidad de crecimiento» dividiendo cada diámetro por la correspondiente edad del árbol.
- Realice un diagrama de dispersión de $Y =$ edad frente a $X =$ velocidad de crecimiento y ajuste una recta de regresión a los datos.
- Dibuje un gráfico de residuos de la regresión del apartado (b). Realice después una gráfica de probabilidad normal de los residuos. ¿Qué dice esa gráfica sobre la cuestión del uso de un modelo lineal y procedimientos de inferencia basados en la regresión?

- Tome el logaritmo de cada valor de la edad y de cada valor de velocidad de crecimiento. Realice un diagrama de dispersión de $Y = \log(\text{edad})$ frente a $X = \log(\text{velocidad de crecimiento})$ y ajuste una recta de regresión a los datos.
- Realice una gráfica de residuos de la regresión del apartado (d). Dibuje después una gráfica de probabilidad normal de los residuos. Basándose en esta gráfica, ¿parece apropiado el modelo de regresión en escala logarítmica del apartado (d)?

12.S.18 (Ejercicio para computador) Unos investigadores midieron las presiones sanguíneas de 22 estudiantes en dos situaciones: cuando los estudiantes estaban relajados y cuando los estudiantes estaban realizando un examen importante. La tabla muestra las presiones sistólica y diastólica de cada estudiante en cada situación³⁷.

Durante el examen		Relajados	
Presión sistólica (mmHg)	Presión diastólica (mmHg)	Presión sistólica (mmHg)	Presión diastólica (mmHg)
132	75	110	70
124	170	90	75
110	65	90	65
110	65	110	80
125	65	100	55
105	70	90	60
120	70	120	80
125	80	110	60
135	80	110	70
105	80	110	70
110	70	85	65
110	70	100	60
110	70	120	80
130	75	105	75
130	70	110	70
130	70	120	80
120	75	95	60
130	70	110	65
120	70	100	65
120	80	95	65
120	70	90	60
130	80	120	70

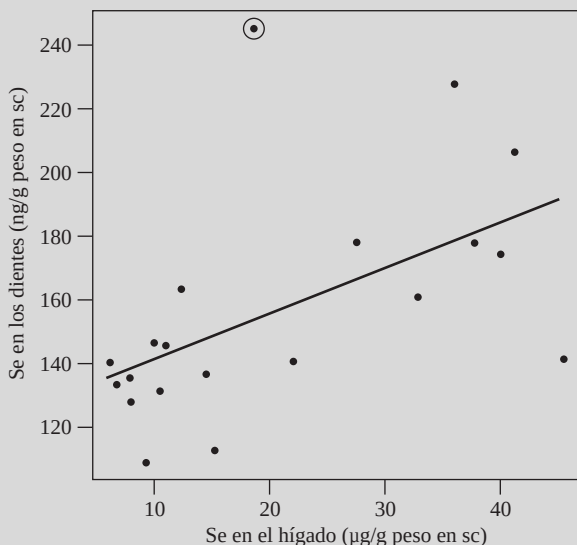
- Calcule el cambio en la presión sistólica restando la presión sistólica en el caso relajado de la presión sistólica durante el examen. Denomine a esta variable X .

- (b) Repita el apartado (a) para la presión diastólica. Denomine la variable resultante Y .
- (c) Realice un diagrama de dispersión de Y con respecto a X y ajuste una recta de regresión a los datos.
- (d) Realice un diagrama de residuos de la regresión del apartado (c).
- (e) Observe el *outlier* en el diagrama de residuos [y en el diagrama de dispersión del apartado (c)]. Elimine el *outlier* del conjunto de datos. Repita después los apartados (c) y (d).
- (f) ¿Cuál es el modelo de regresión ajustado (después de haber eliminado el *outlier*)?

12.S.19 (Continuación del Ejercicio 12.S.18) Considere los datos del Ejercicio 12.S.18, apartado (f).

- (a) Construya un intervalo de confianza del 95 % para β_1 .
- (b) Interprete el intervalo de confianza del apartado (a) en el contexto de este ejercicio.

12.S.20 El selenio (Se) es un elemento esencial que ha demostrado jugar un papel importante en la protección de los mamíferos marinos contra los efectos tóxicos del mercurio (Hg) y otros metales. Se ha sugerido que las concentraciones de metales en los dientes de los mamíferos marinos se pueden utilizar potencialmente como indicadores de la carga corporal. Se recogieron 20 ballenas (*Delphinapterus leucas*) en el Mackenzie Delta, Northwest Territories, en 1996, como parte de la caza tradicional anual de los Inuit. Las concentraciones de selenio en los dientes y en el hígado se resumen en la tabla y se representan a continuación³⁸.



- (a) ¿Se puede considerar la correlación muestral entre el selenio en dientes (Y) e hígado (X), $r = 0,53726$, como una estimación del coeficiente de correlación poblacional? Explique brevemente su respuesta.

Ballena	Se en el hígado ($\mu\text{g/g}$)	Se en los dientes (ng/g)	Ballena	Se en el hígado ($\mu\text{g/g}$)	Se en los dientes (ng/g)
1	6,23	140,16	11	15,28	112,63
2	6,79	133,32	12	18,68	245,07
3	7,92	135,34	13	22,08	140,48
4	8,02	127,82	14	27,55	177,93
5	9,34	108,67	15	32,83	160,73
6	10,00	146,22	16	36,04	227,60
7	10,57	131,18	17	37,74	177,69
8	11,04	145,51	18	40,00	174,23
9	12,36	163,24	19	41,23	206,30
10	14,53	136,55	20	45,47	141,31

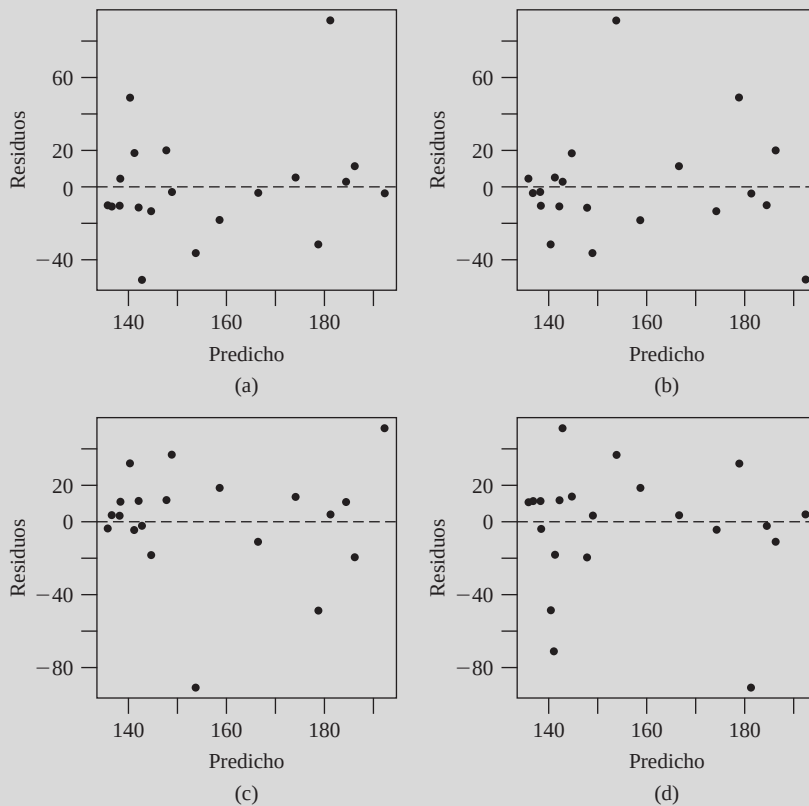
- (b) Si se eliminara del conjunto de datos el punto rodeado con un círculo, el valor de correlación muestral dado en el apartado (a), ¿aumentaría, disminuiría o permanecería constante?
- (c) Si se invirtieran los papeles de X e Y (es decir, Y = selenio en el hígado y X = selenio en los dientes), el valor de correlación muestral dado en el apartado (a), ¿aumentaría, disminuiría o permanecería constante?
- (d) ¿Es el punto de la gráfica rodeado con un círculo un punto de apalancamiento y/o de influencia? Explique brevemente su respuesta.
- (e) ¿Es el punto de la gráfica rodeado con un círculo un *outlier*?

12.S.21 (Continuación del 12.S.20) Se presentan a continuación estadísticos resumen para los datos del selenio en el Ejercicio 12.S.20.

$$\begin{aligned}\bar{x} &= 20,684 & \bar{y} &= 156,599 \\ s_x &= 13,4489 & s_y &= 36,0586 \\ r &= 0,53726 & \text{SC(resid)} &= 17.573,3\end{aligned}$$

- (a) Calcule la recta de regresión del selenio en los dientes con respecto a selenio en el hígado.
- (b) Calcule un intervalo de confianza del 95 % para la pendiente de la recta de regresión.
- (c) Interprete el intervalo calculado en el apartado (b) en el contexto del problema.
- (d) Utilizando el intervalo calculado en el apartado (b), ¿es razonable creer que la pendiente es tan pequeña como 0,25 (ng/g)/(μg/g)?

12.S.22 (Continuación de los Ejercicios 12.S.20 y 12.S.21) En referencia a los datos representados en el Ejercicio 12.S.20, ¿cuál de las gráficas siguientes es una gráfica de residuos resultante del ajuste de la recta de regresión en el Ejercicio 12.S.21, apartado (a)? Justifique su elección.



12.S.23 (Continuación del Ejercicio 12.S.20) Las ballenas observadas en este estudio se cazaron durante la caza tradicional Inuit en dos años particulares. ¿Qué estamos suponiendo sobre las ballenas capturadas para justificar nuestro análisis de estos datos en los problemas anteriores?

Notas

1. Datos no publicados cortesía de M. B. Nichols y R. P. Maickel. El experimento original contenía más de tres grupos de tratamiento.
2. Bodgan, K. y Schenk, M. (2009). Evaluation of soil characteristics potentially affecting arsenic concentration in paddy rice (*Oryza sativa* L.). *Environmental Pollution* **157**, 2617-2621. Datos de 2006 digitalizados de la Figura 3.
3. Adaptado de Andren, C. y Nilson, G. (1981). Reproductive success and risk of predation in normal and melanistic colour morphs of the adder, *Vipera berus*. *Biological Journal of the Linnean Society* **15**, 235-246. (Los datos son de las mujeres melanísticas; los valores se han modificado ligeramente para simplificar la exposición).
4. Erne, P., Bolli, P., Buerigisser, E. y Buehler, F. R. (1984). Correlation of platelet calcium with blood pressure. *New*

England Journal of Medicine **310**, 1084-1088. Reimpreso con permiso. Datos iniciales cortesía de F. R. Buehler. Para simplificar la explicación, hemos omitido nueve pacientes con presión sanguínea «en la frontera».

5. Adaptado de Spencer, D. F., Volpp, T. R. y Lembi, C. A. (1980). Environmental control of *Pithophora oedogonia* (Chlorophyceae) akinete germination. *Journal of Phycology* **16**, 424-427. El valor $r = -0,72$ se ha calculado a partir de datos representados gráficamente por Spencer *et al.*

6. Albert, A. (1981). Atypicality indices as reference values for laboratory data. *American Journal of Clinical Pathology* **76**, 421-425.

7. Harding, A. J., Wong, A., Svoboda, M., Kril, J. J. y Halliday, G. M. (1997). Chronic alcohol consumption does

not cause hippocampal neuron loss in humans. *Hippocampus* **7**, 78-87. El valor $r = -0,63$ se ha calculado a partir de datos representados gráficamente por Harding *et al.*

8. Smith, R. D. (1978-1979). Institute of Agricultural Engineering Annual Report. Salisbury, Zimbabwe: Department of Research and Specialist Services, Ministry of Agriculture. Datos iniciales cortesía de R. D. Smith.

9. Bowers, W. S., Hoch, H. C., Evans, P. H. y Katayama, M. (1986). Thallophtytic allelopathy: Isolation and identification of laetisarinic acid. *Science* **232**, 105-106. Copyright 1986 de la AAAS. Datos iniciales cortesía de los autores.

10. Webb, P. (1981). Energy expenditure and fat-free mass in men and women. *American Journal of Clinical Nutrition* **34**, 1816-1826.

11. Huel, G., *et al.* (2008). Hair mercury negatively correlates with calcium pump activity in human term newborns and their mothers at delivery. *Environmental Health Perspectives* **116**, 263-267.

12. Datos simulados basados en temperaturas oceánicas típicas cerca de Morro Bay, California.

13. Cicirelli, M. F., Robinson, K. R. y Smith, L. D. (1983). Internal pH of *Xenopus* oocytes: A study of the mechanism and role of pH changes during meiotic maturation. *Developmental Biology* **100**, 133-146. Datos iniciales cortesía de M. F. Cicirelli.

14. Maickel, R. P. y Nash, J. F., Jr. (1985). Differing effects of short-chain alcohols on body temperature and coordinated muscular activity in mice. *Neuropharmacology* **24**, 83-89. Reimpreso con permiso. Copyright 1985, Pergamon Journals, Ltd. Datos iniciales cortesía de J. F. Nash, Jr.

15. Adaptado de Barclay, A. M. y Crawford, R. M. M. (1984). Seedling emergence in the rowan (*Sorbus aucuparia*) from an altitudinal gradient. *Journal of Ecology* **72**, 627-636. Reimpreso con permiso de John Wiley.

16. Olson, J. M. y Mardh, R. L. (1998). Activation patterns and length changes in hindlimb muscles of the bullfrog *Rana catesbeiana* during jumping. *The Journal of Experimental Biology* **201**, 2763-2777.

17. Sulcove, J. A. y Lacuesta, N. N. (1998). The effect of gender and height on peak flow rate. Manuscrito no publicado, Oberlin College.

18. Hamill, P. V. V., Johnston, F. E. y Lemeshow, S. (1973). Height and weight of youths 12-17 years, United States. *U.S. National Center for Health Statistics, Vital and Health Statistics*, Serie 11, n.º 124. Washington, D.C.: U.S. Department of Health, Education and Welfare. Las distribuciones condicionales de peso dada la altura se representan en la Figura 12.4.1 como distribuciones normales. La población ficticia se ajusta bien a la población real (descrita por Hamill *et al.*) en la parte central de cada distribución condicional, pero las distribuciones condicionales reales tienen colas izquierdas más cortas y

colas derechas más largas que las distribuciones condicionales normales ficticias.

19. Maickel, R. P. Comunicación personal.

20. Datos ficticios pero realistas, basados en variación inter e intraindividuos descrita en Williams, G. Z., Widdowson, G. M. y Penton, J. (1978). Individual character of variation in time-series studies of healthy people. II. Differences in values for clinical chemical analytes in serum among demographic groups, by age and sex. *Clinical Chemistry* **24**, 313-320.

21. Stewart, T. S., Nelson, L. A., Perry, T. W. y Martin, T. G. (1985). Datos no publicados cortesía de T. S. Stewart.

22. Pappas, T. y Mitchell, C. A. (1985). Effects of seismic stress on the vegetative growth of *Glycine max* (L.) Merr. cv. Wells II. *Plant, Cell and Environment* **8**, 143-148. Reimpreso con permiso de John Wiley. Datos iniciales cortesía de los autores.

23. Fialho, E. T., Ferreira, A. S., Freitas, A. R. y Albino, L. F. T. (1982). Energy and nitrogen balance of ration (corn-soybean meal) for male castrated and non-castrated swine of different weights and breeds (in Portuguese). *Revista Sociedade Brasileira de Zootecnia* **11**, 405-419. Datos iniciales cortesía de E. T. Fialho.

24. Ejemplo comunicado por D. A. Holt.

25. Chambers, J. Q., Higuhi, N. y Schimel, J. (1998). Ancient trees in Amazonia. *Nature* **391**, 135-136. Datos iniciales cortesía de J. Chambers.

26. Dikmen, S. y Hansen, P. (2009). Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? *Journal of Dairy Science* **92**, 109-116. Los datos se simularon para producir resultados similares a los presentados en el artículo.

27. Florey, C. du V. y Acheson, R. M. (1969). Blood pressure as it relates to physique, blood glucose, and serum cholesterol. *U.S. National Center for Health Statistics*, Serie 11, n.º 34. Washington, D.C.: U.S. Department of Health, Education and Welfare.

28. Bernays, E. A. (1986). Diet-induced head allometry among foliage-chewing insects and its importance for graminivores. *Science* **231**, 495-497. Copyright 1986 de la AAAS. Datos iniciales cortesía del autor.

29. Hibi, K., Taguchi, M., Nakayama, H., Takase, T., Kasai, Y., Ito, K., Akiyama, S. y Nakao, A. (2001). Molecular detection of p16 promoter methylation in the serum of patients with esophageal squamous cell carcinoma. *Clinical Cancer Research* **7**, 3135-3138. Había 38 pacientes en el estudio; solo los 31 en los que «el ADN del tumor estaba metilado» se incluyen en este análisis.

30. Gwynne, D. T. (1981). Sexual difference theory: Mormon crickets show role reversal in mate choice. *Science* **213**, 779-780. Copyright 1981 by the AAAS. Cálculos

basados en los datos iniciales proporcionados por cortesía del autor.

31. Heggestad, H. E. y Bennett, J. H. (1981). Photochemical oxidants potentiate yield losses in snap beans attributable to sulfur dioxide. *Science* **213**, 1008-1010. Copyright 1981 de la AAAS. Datos iniciales cortesía de H. E. Heggestad.

32. Thirakhupt, K. (1985). Foraging ecology of sympatric parids: Individual and population responses to winter food scarcity. Ph. D. thesis, Purdue University. Datos iniciales cortesía del autor y de K. N. Rabenold.

33. Datos no publicados cortesía de A. H. Ismail y L. S. Verity.

34. Tazawa, H., Pearson, J. T., Komoro, T. y Ar, A. (2001). Allometric relationships between embryonic heart rate and fresh egg mass in birds. *The Journal of Experimental Biology* **204**, 165-174.

35. Conjuntos de datos inventados por F. J. Anscombe. Véase Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician* **27**, 17-21.

36. Datos no publicados cortesía de M. B. Nichols y R. P. Maickel. El experimento contenía realmente más de tres grupos. Los datos del Ejemplo 12.1.1 son de otra parte del estudio, usando una forma química diferente de anfetamina.

37. Marazziti, D., DiMuro, A. y Castrogiovanni, P. (1992). Psychological stress and body temperature changes in humans. *Psychology & Behavior* **52**, 393-395.

38. Kinghorn, A., Humphries, M., Outridge, P. y Chan, H. M. (2008). Teeth as biomonitors of selenium concentrations in tissues of beluga whales (*Delphinapterus leucas*). *Science of the Total Environment* **402**, 43-50. Datos digitalizados de la Figura 3.

RESUMEN DE MÉTODOS DE INFERENCIA

Objetivos

En este capítulo resumiremos los métodos de inferencia presentados en el texto:

- mostraremos el proceso para escoger una técnica de inferencia apropiada entre las presentadas en los capítulos anteriores;
- consideraremos varios ejemplos de selección de métodos de inferencia.

13.1 Introducción

En los Capítulos 2 y 6 al 12 hemos presentado muchos métodos estadísticos para resumir datos de forma visual y numérica y para realizar inferencias. Los estudiantes de estadística quedan a menudo abrumados por el número y la variedad de procedimientos presentados. Lo que un estadístico ve como un conjunto de herramientas claramente dispuestas para analizar datos, puede parecer borroso al estudiante novel. En este capítulo presentamos una variedad de ejemplos que demuestran los procesos de análisis desde la exploración y el resumen hasta la inferencia, utilizando algunos de los métodos presentados en capítulos anteriores. Con los ejemplos, proporcionamos también algunas pautas que son útiles para decidir cómo realizar una inferencia a partir de un conjunto dado de datos.

Cuando nos enfrentamos con un conjunto de datos, es útil preguntarse una serie de cuestiones:

1. *¿Qué cuestión intentaban responder los investigadores cuando recogieron los datos?* El análisis de datos se realiza con un propósito: extraer información y tomar decisiones. Cuando se observan los datos, es de ayuda tener en mente el propósito para el que se tomaron dichos datos. Por ejemplo, ¿deseaban los investigadores comparar grupos, quizá pacientes que recibieron un nuevo medicamento con pacientes que recibieron un placebo? ¿Intentaban ver la relación entre dos variables cuantitativas, de forma que con una variable se pudieran realizar predicciones sobre la otra? ¿Estaban comprobando si un modelo hipotético proporcionaba una predicción exacta de las probabilidades asociadas a una variable categórica? Una buena comprensión de por qué se tomaron los datos a menudo clarifica la cuestión siguiente:
2. *¿Cuál es la variable de respuesta del estudio?* Por ejemplo, si los investigadores estaban considerando el efecto de una medicación en la presión sanguínea, entonces probablemente la variable de respuesta es Y = cambio en la presión sanguínea de un individuo (una variable numérica continua). Si estaban considerando si una medicación cura o no cura una enfermedad, entonces la variable de respuesta es categórica con dos niveles: sí, si la persona se cura y no, si la persona no se cura, o incluso podría ser categórica con tres o más niveles ordenados: completamente curado, mejora, no mejora.
3. *¿Qué variables de predicción, si existen, intervienen?* Por ejemplo, si se está comparando un nuevo medicamento con un placebo, entonces la variable de predicción es la pertenencia al grupo: un paciente pertenece al grupo que recibe el nuevo medicamento o pertenece al grupo del placebo. Si se usa la altura para predecir el peso, entonces la

altura es la variable de predicción (y el peso es la variable de respuesta). Algunas veces no existe la variable de predicción. Por ejemplo, un investigador podría estar interesado en la distribución de los niveles de colesterol en adultos. En este caso, la variable de respuesta es el nivel de colesterol, pero no hay variable de predicción. (Se podría argumentar que existe un predictor: si alguien es o no adulto. Si deseáramos comparar los niveles de colesterol en adultos con los de niños, entonces si se es o no adulto sería un predictor. Pero si no se realiza ninguna comparación, de forma que todas las personas del estudio forman parte del mismo grupo (adultos), entonces no es adecuado hablar de una variable de *predicción*, ya que la pertenencia a un grupo no varía de una persona a otra).

Las respuestas a esas preguntas ayudan a encuadrar el análisis a realizar. Algunas veces el análisis será completamente descriptivo y no incluirá ninguna inferencia estadística, por ejemplo en el caso en el que los datos no se recogen mediante muestreo aleatorio. En los casos en los que hay que realizar una inferencia estadística, en general existe más de una forma de proceder. Dos estadísticos que analicen el mismo conjunto de datos podrían utilizar métodos algo diferentes y extraer conclusiones distintas. Sin embargo, hay procedimientos estadísticos de uso común en varias situaciones. El organigrama que se presenta en la Figura 13.1.1 ayuda a organizar los métodos de inferencia que se han presentado en este libro.

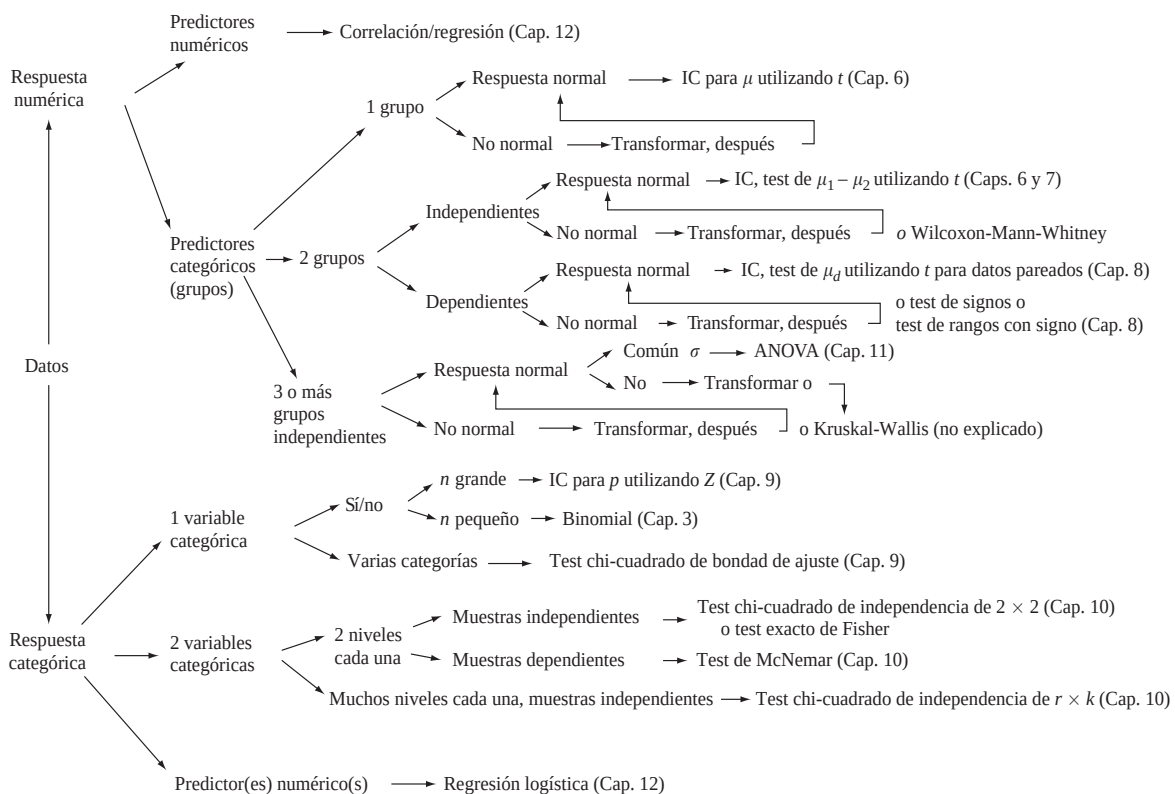


Figura 13.1.1 Organigrama de métodos de inferencia

Para utilizar este organigrama, empezamos preguntándonos si la variable de respuesta es cuantitativa o categórica. Después consideramos el tipo de variables de predicción del estudio y si las muestras tomadas son independientes entre sí o son dependientes (por ejemplo, parejas ajustadas). Muchos métodos, como el intervalo de confianza para la media poblacional presentado en el Capítulo 6 dependen de que los datos procedan de una población que tenga distribución normal (esta condición es menos importante para muestras grandes que para muestras pequeñas, debido al Teorema Central del Límite). Los datos que no son normales se pueden transformar a menudo para aproximarse a la normalidad, y aplicar después métodos basados en distribución normal. Si las transformaciones fallan en conseguir normalidad aproxi-

mada, entonces se pueden usar métodos no paramétricos como el test de Wilcoxon-Mann-Whitney o el test de rangos con signo de Wilcoxon.

Nótese que el organigrama solo dirige la atención a la colección de métodos presentados en los capítulos anteriores. Cuidado con la falacia de Mark Twain: «Cuando tu única herramienta es un martillo, todos los problemas parecen clavos». No todos los problemas de inferencia estadística se pueden resolver con los métodos presentados aquí. En particular, estos métodos se centran en la consideración de parámetros, como la media poblacional, μ , o la proporción poblacional, p . Algunas veces los investigadores están interesados en otros aspectos de las distribuciones, como el percentil 75. Si tenemos dudas sobre cómo proceder en un análisis, es mejor consultar a un estadístico.

Análisis exploratorio de datos

Independientemente del tipo de análisis que estemos considerando, siempre es buena idea comenzar representando una o varias gráficas de los datos. El tipo de gráficas depende del tipo de datos que se están analizando. Por ejemplo, cuando se comparan dos muestras de datos cuantitativos, diagramas de puntos o de caja uno al lado del otro son informativos, tanto para realizar una comparación visual de las dos muestras como para evaluar si los datos satisfacen la condición de normalidad. Al analizar datos categóricos, los diagramas de barras son de utilidad. Al tratar con dos variables cuantitativas, son de utilidad los diagramas de dispersión.

Hay que tener en mente que el análisis estadístico intenta ayudar a comprender el problema científico que se tiene entre manos. Por tanto, las conclusiones se deben plantear en el contexto del estudio científico. En la Sección 13.2 presentamos algunos ejemplos de conjuntos de datos y los tipos de análisis que se podrían realizar con ellos.

13.2 Ejemplos de análisis de datos

En esta sección consideramos varios conjuntos de datos y los tipos de análisis que son apropiados para los mismos. Las tres preguntas planteadas en la Sección 13.1 y el organigrama de la Figura 13.1. proporcionan un marco para el seguimiento de los ejemplos siguientes.

Ejemplo 13.2.1 Ácido giberélico

Se piensa que el ácido giberélico (AG) aumenta la longitud de los tallos de las plantas. Unos investigadores realizaron un experimento para investigar el efecto del AG en una variedad mutante del género *Brassica* denominada *ros*. Aplicaron AG a 17 plantas y aplicaron agua a 15 plantas de control. Tras 14 días midieron el crecimiento de las 32 plantas. El crecimiento medio de las 15 plantas de control fue de 26,7 mm, con una DT de 37,5 mm. Para las 17 plantas con AG el crecimiento medio fue de 92,6 mm, con una DT de 41,7 mm. Los datos se muestran en la Tabla 13.2.1 y se representan en la Figura 13.2.1¹.

Volvamos a las tres preguntas planteadas en la Sección 13.1. (1) En este experimento, los investigadores intentaban establecer si el AG afecta a la velocidad de crecimiento del *ros*; (2) la variable de respuesta es el crecimiento de *ros* durante 14 días, que es numérica; (3) la variable de predicción es la pertenencia a un grupo (grupo de AG o grupo de control) y es categórica. Los dos grupos son independientes entre sí.

El organigrama de la Figura 13.1.1 nos dirige a considerar un test t de dos muestras, si los datos son normales o se pueden transformar para que lo sean, o a un test de Wilcoxon-Mann-Whitney. La Figura 13.2.2 muestra que la distribución de la muestra de control es marcadamente no normal, por lo que se requiere una transformación.

Tomando logaritmos de las observaciones resultan los diagramas de puntos y gráficas de probabilidad normal de las Figuras 13.2.3 y 13.2.4.

En escala logarítmica los datos no muestran evidencia de no normalidad (los P valores de la prueba de Shapiro-Wilks para el control y el AG son 0,2083 y 0,2296, respectivamente), por lo que podemos proceder con un test t de dos muestras. Las desviaciones típicas de las dos muestras son bastante diferentes, como podemos ver en la Figura 13.2.3. Sin embargo, un test t no conjunto es aún apropiado. La salida de computador siguiente muestra que

Tabla 13.2.1 Crecimiento de plantas (mm) tras 14 días

	Control	AG
	3	71
	2	87
	34	117
	13	80
	6	112
	118	66
	14	128
	107	153
	30	131
	9	45
	3	38
	3	137
	49	57
	4	163
	6	47
		108
		35
Media	26,7	92,6
DT	37,5	41,7

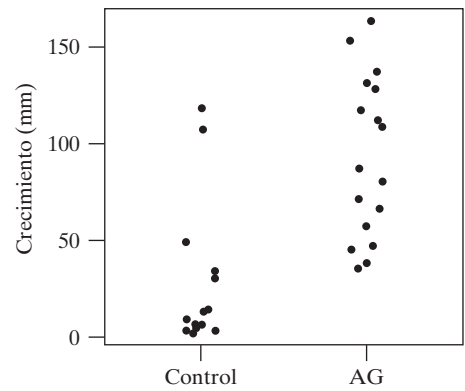


Figura 13.2.1 Diagramas de puntos de crecimiento de plantas ros (mm) tras 14 días

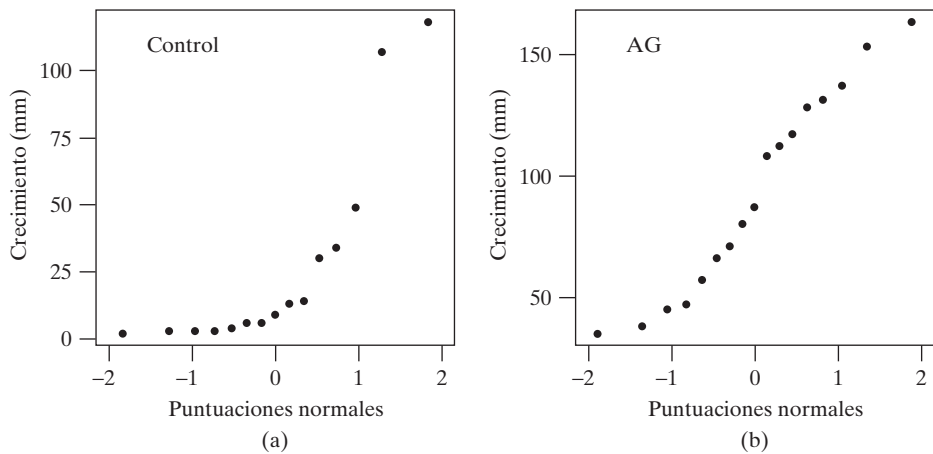


Figura 13.2.2 Gráficas de probabilidad normal de (a) datos de control y (b) datos de AG

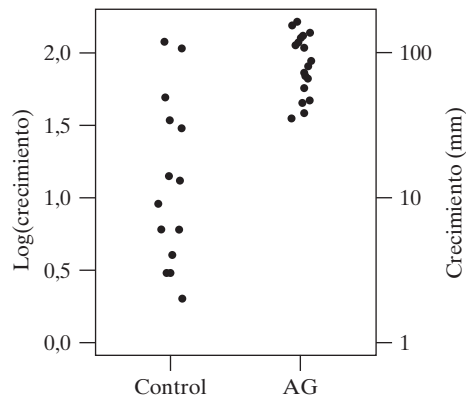


Figura 13.2.3 Diagramas de puntos de log(crecimiento) de plantas *ros* (mm) tras 14 días

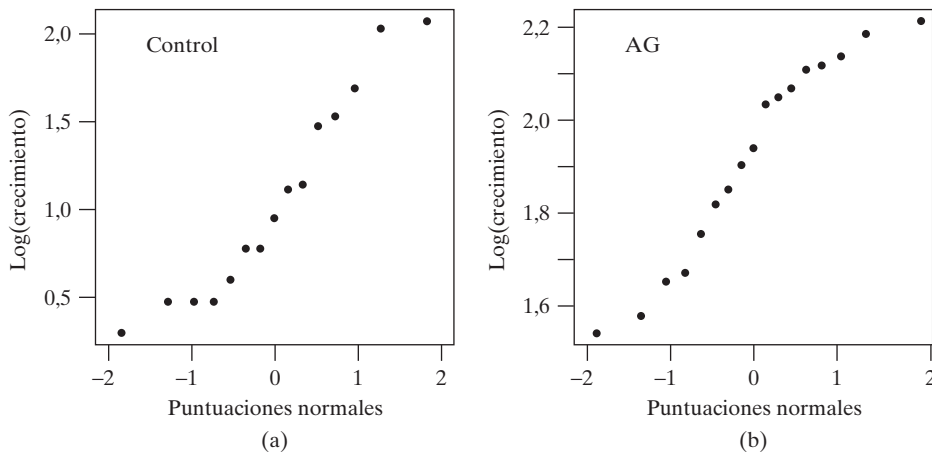


Figura 13.2.4 Gráficas de probabilidad normal de (a) datos de control y (b) datos de AG en escala logarítmica

$t_s = -5,392$ y el P valor es muy pequeño. Por tanto, tenemos evidencia fuerte de que el AG aumenta el crecimiento del *ros*. ■

Two Sample t-test

data: log10(Crecimiento)

$t = -5.3917$, $df = 17.445$, $p\text{-value} < 0.0001$

alt. hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.1943596, -0.5234687

Ejemplo 13.2.2 Velocidad de nado de ballenas

Un biólogo estaba interesado en la relación entre la velocidad de nado de una ballena beluga y su frecuencia de batido de cola. Se estudió una muestra de 19 ballenas y se tomaron medidas de su velocidad de nado, medida en unidades de longitud corporal de la ballena por segundo (de manera que un valor de 1,0 indica que la ballena se mueve hacia delante una longitud corporal, L , por segundo) y de frecuencia de batido de cola, medida en hercios (de forma que un valor de 1,0 significa un batido de cola por segundo)². Los datos son los siguientes:

Ballena	Velocidad (L/s)	Frecuencia (Hz)	Ballena	Velocidad (L/s)	Frecuencia (Hz)
1	0,37	0,62	11	0,68	1,20
2	0,50	0,675	12	0,86	1,38
3	0,35	0,68	13	0,68	1,41
4	0,34	0,71	14	0,73	1,44
5	0,46	0,80	15	0,95	1,49
6	0,44	0,88	16	0,79	1,50
7	0,51	0,88	17	0,84	1,50
8	0,68	0,92	18	1,06	1,56
9	0,51	1,08	19	1,04	1,67
10	0,67	1,14			

Sería natural preguntarse: «¿Cuando la cola bate más fuerte, la ballena se mueve más rápido?», pero el biólogo que realizaba la prueba se centró en la cuestión relacionada: «¿Depende la frecuencia de batido de la cola de la velocidad?». Para la cuestión del biólogo, la variable de respuesta, la frecuencia, es numérica, y la variable de predicción es la velocidad, que es también numérica. Por tanto, podemos considerar el uso del análisis de regresión para estudiar la relación entre la velocidad y la frecuencia. La Figura 13.2.5 es un diagrama de dispersión de los datos, que muestra una tendencia creciente en la frecuencia a medida que la velocidad aumenta.

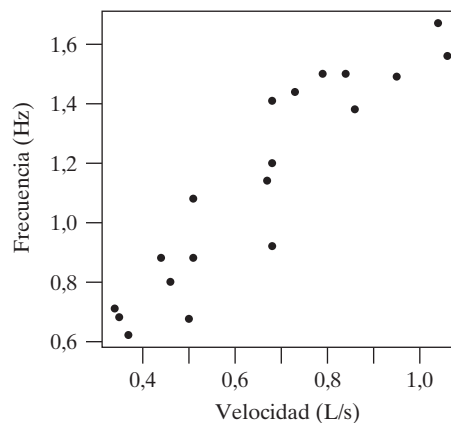


Figura 13.2.5 Diagrama de dispersión de la frecuencia frente a la velocidad

Un modelo de regresión para estos datos es $Y = \beta_0 + \beta_1 X + \varepsilon$. Ajustando el modelo a los datos se llega a la ecuación $\hat{y} = 0,19 + 1,439x$, o Frecuencia = $0,19 + 1,439 \times$ Velocidad, como muestra la salida de computador siguiente. La Figura 13.2.6 muestra la gráfica de residuos de este ajuste. El hecho de que esta gráfica no muestre patrones claros da fundamento al uso del modelo de regresión.

Coefficients:

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.1895    0.1004    1.887  0.0763
Velocity       1.4393    0.1451    9.917  1.75e-08
Residual standard error: 0.1396 on 17 degrees of freedom
R-squared: 0.8526

```

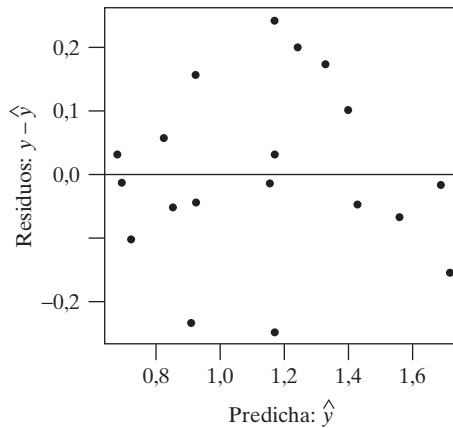


Figura 13.2.6 Gráfica de residuos para el ajuste por regresión de la frecuencia

La hipótesis nula

$$H_0: \beta_1 = 0$$

se contrasta con un test t , como se muestra en la salida de la regresión. La gráfica de probabilidad normal de los residuos, que se muestra en la Figura 13.2.7, fundamenta el uso del test t en este caso, ya que indica que la distribución de los 19 residuos es consistente con lo que esperaríamos ver si los errores aleatorios provinieran de una distribución normal. El estadístico t tiene 17 grados de libertad y un P valor que es menor que 0,0001. Por tanto, la evidencia de que la frecuencia está relacionada con la velocidad es muy fuerte. Rechazamos la afirmación de que la tendencia lineal de los datos se debe al azar.

Continuando con el análisis, la salida de computador muestra que r^2 es 85,3 %. Por tanto, en la muestra el 85,3 % de la variabilidad de la frecuencia se puede explicar por la variabilidad de la velocidad. (Esto es significativamente diferente de cero, como indica el test t de $H_0: \beta_1 = 0$). ■

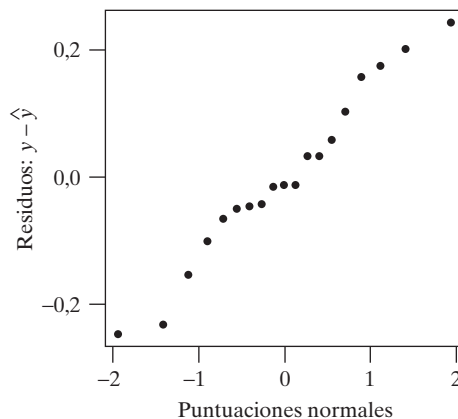


Figura 13.2.7 Gráfica de probabilidad normal de los residuos para el ajuste por regresión de la frecuencia

Ejemplo 13.2.3 Tamoxifeno

En un experimento aleatorizado y doble ciego, se suministró el medicamento tamoxifeno a 6.681 mujeres y a otras 6.707 se les suministró un placebo. Después de 4 años se produjeron 89 casos de cáncer de pecho en el grupo del tamoxifeno, y 175 en el grupo del placebo³.

El propósito de este experimento fue determinar si el tamoxifeno es efectivo en la prevención del cáncer. Nótese que como se trata de un experimento, y no de un estudio observacional, podemos hablar en términos de relación causa-efecto. La variable de respuesta es si una mujer desarrolló o no desarrolló cáncer. La variable de predicción es la pertenencia a un grupo (es decir, si la mujer recibió o no tamoxifeno). La Figura 13.2.8 es un diagrama de barras de los datos, que muestra que el cáncer fue mucho más común en el grupo del placebo.

Los datos se organizaron en una tabla de contingencia de 2×2 , como la Tabla 13.2.2. Un test chi-cuadrado de independencia dio un resultado de $\chi_s^2 = 28,2$. Con 1 grado de libertad, el P valor de este test es casi cero. Hay una evidencia muy fuerte a favor de que el tamoxifeno reduce la probabilidad de cáncer de pecho.

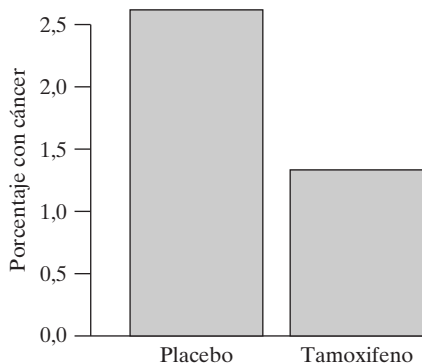


Figura 13.2.8 Diagrama de barras de los datos del tamoxifeno

Tabla 13.2.2 Datos de tamoxifeno			
	Tratamiento		
	Placebo	Tamoxifeno	
Cáncer	175	89	264
No cáncer	6.532	6.592	13.124
Total	6.707	6.681	13.388

Podemos también construir un intervalo de confianza con estos datos. De los pacientes del placebo, $\frac{175}{6.707}$ o el 2,61 % desarrolló cáncer, de forma que $\tilde{p}_1 = \frac{175 + 1}{6.707 + 2} = 0,0262$. De los pacientes del tamoxifeno, $\frac{89}{6.681}$ o el 1,33 % desarrolló cáncer, de forma que $\tilde{p}_2 = \frac{89 + 1}{6.681 + 2} = 0,0135$. El error típico de la diferencia es

$$\begin{aligned} ET_{(\tilde{p}_1 - \tilde{p}_2)} &= \sqrt{\frac{(0,0262)(1 - 0,0262)}{6.707 + 2} + \frac{0,0135(1 - 0,0135)}{6.681 + 2}} \\ &= 0,0024 \end{aligned}$$

Un intervalo de confianza del 95 % para $p_1 - p_2$ es $(0,0262 - 0,0135) \pm 1,96(0,0024)$ o $(0,0080, 0,0174)$. Por tanto, tenemos un 95 % de confianza en que el tamoxifeno reduce la probabilidad de cáncer de pecho entre 0,80 y 1,74 puntos porcentuales.

Podemos calcular también el riesgo relativo de cáncer. El riesgo relativo estimado es

$$\frac{\Pr\{\text{Cáncer} | \text{Placebo}\}}{\Pr\{\text{Cáncer} | \text{Tamoxifeno}\}} = \frac{0,0261}{0,0133} = 1,96$$

Por consiguiente, estimamos que el cáncer de pecho es 1,96 veces más probable tomando placebo que tomando tamoxifeno. ■

Ejemplo 13.2.4 Abultamientos cromosómicos

Las proteínas de choque térmico (PCT) son un tipo de proteínas producidas por algunos organismos para protegerse del daño causado por la exposición a altas temperaturas. En la mosca de la fruta (*Drosophila melanogaster*), los genes que codifican las PCT se encuentran en cromosomas que se desenrollan y parecen hincharse. Este abultamiento cromosómico se puede ver al microscopio. Un biólogo contó el número de abultamientos por brazo cromosómico

en las glándulas salivales de 40 larvas de *Drosophila* que se habían sometido a una temperatura de 37 °C durante 30 minutos, 40 larvas que se habían sometido a una temperatura de 37 °C durante 60 minutos, y 40 larvas de control.

El propósito de este experimento era determinar el efecto, si lo había, del choque térmico en las PCT. La variable de respuesta es el número de abultamientos en un brazo cromosómico, que es numérica. La variable de predicción, la pertenencia a un grupo (control, 30 minutos o 60 minutos), es categórica. La Figura 13.2.9 presenta los diagramas de puntos de los datos. Dichos datos se resumen en la Tabla 13.2.3⁴.

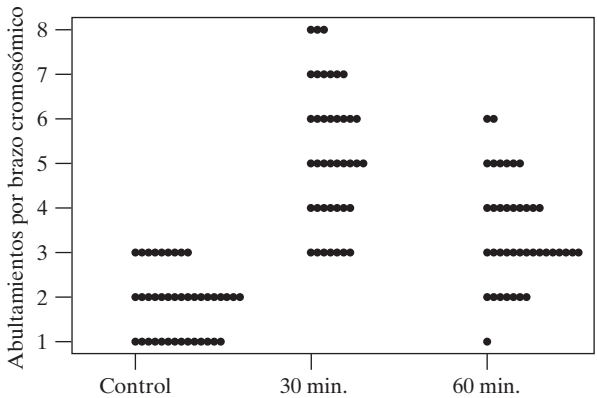


Figura 13.2.9 Diagramas de puntos de los abultamientos por brazo cromosómico en el experimento de choque térmico con *Drosophila*

Tabla 13.2.3 Abultamientos por brazo cromosómico en el experimento de choque térmico con <i>Drosophila</i>			
Grupo	<i>n</i>	Media	DT
Control	40	1,88	0,76
30 min.	40	5,20	1,54
60 min.	40	3,45	1,18

Los diagramas de puntos sugieren un efecto debido al choque térmico (y podemos hablar de un efecto, no solo de una asociación, porque se trata de un experimento). Esta impresión visual se puede confirmar con un análisis de varianza. Las gráficas muestran también que la distribución toma solo unos pocos valores en cada caso, por lo que la condición de normalidad para el ANOVA no se cumple. Sin embargo, ya que las distribuciones parece ser razonablemente simétricas, los tamaños muestrales moderadamente grandes e iguales y las DT similares entre los grupos, podemos tener confianza en el *P* valor del ANOVA. La siguiente salida de computador del ANOVA confirma que hay una fuerte evidencia en contra de $H_0: \mu_1 = \mu_2 = \mu_3$. Concluimos que el choque térmico incrementa, de hecho, el número de abultamientos por brazo cromosómico.

	Df	Sum Sq	Mean Sq	F value	Pr (>F)
Group	2	221.32	110.658	76.757	<0.0001
Residuals	117	168.68	1.442		
Total	119	390.00			

Como una extensión del ANOVA, podríamos considerar un contraste que comparara la media del grupo control con el promedio de las dos medias de los grupos de choque térmico. ■

Ejemplo 13.2.5 Toque terapéutico

El toque terapéutico (TT) es una forma de medicina alternativa en la que un profesional manipula el campo de energía humana del paciente. Sin embargo, muchas personas han cuestionado la posibilidad de que los profesionales del TT detecten el campo de energía humana, e incluso si dicho campo existe. Un investigador probó la capacidad de 28 profesionales de TT como sigue. Se puso una pantalla entre el investigador y el profesional, que se sentaron en los extremos opuestos de una mesa. El profesional extendió sus manos bajo la pantalla y las puso sobre la mesa, con las palmas boca arriba. El investigador lanzó una moneda para escoger una de las manos del profesional. Después, el investigador mantuvo su mano derecha, con la palma boca abajo, sobre la mano elegida del profesional. Entonces, se pidió al profesional que identificara qué mano se había elegido, como una prueba de si el profesional podía detectar un campo de energía humana proveniente de la mano del investigador.

La prueba se repitió 10 veces con cada uno de los 28 profesionales de TT. El número de detecciones correctas, «éxitos», en las 10 pruebas varió desde 1 hasta 8, con un promedio de 4,4. Hubo 123 éxitos en las 280 pruebas realizadas⁵. La Tabla 13.2.4 muestra la distribución de éxitos entre los 28 profesionales que participaron en la prueba.

Tabla 13.2.4 Distribución de éxitos en 10 pruebas del experimento de toque terapéutico	
Número de éxitos	Número de profesionales
0	0
1	1
2	1
3	8
4	5
5	7
6	2
7	3
8	1
9	0
10	0
Total	28

El objetivo de este experimento era determinar la capacidad de los profesionales de TT para detectar el campo de energía humana. La variable de respuesta es una variable sí/no (categórica): sí, para un éxito, y no, para un fracaso. No hay variable de predicción en este caso, ya que la prueba trabaja con un único grupo de 28 profesionales de TT.

Sea p la probabilidad de un éxito en una de las pruebas del experimento. La hipótesis nula natural es $H_0: p = 0,5$. Una forma de analizar los datos sería realizar un test chi-cuadrado de bondad de ajuste utilizando el total de las 280 pruebas, con una alternativa direccional $H_A: p > 0,5$. El P valor de este contraste es mayor que 0,50, ya que los datos no se desvían de H_0 en la dirección especificada por H_A .

Se podría argumentar que p podría ser mayor que 0,5 para algunos profesionales de TT, pero quizá no para todos ellos. Si p no es la misma para cada uno de los profesionales de TT (tanto si p es 0,5 para alguno de ellos como si no), entonces el test chi-cuadrado de bondad de ajuste utilizando el total de las 280 pruebas no es apropiado, ya que las 280 pruebas no son independientes entre sí. Sin embargo, se podrían analizar los datos de forma separada para cada uno de los 28 profesionales. En este análisis se podría utilizar un modelo binomial, ya que el tamaño de la muestra de $n = 10$ es pequeño. Las probabilidades binomiales se muestran en la Tabla 13.2.5. La probabilidad de 8 o más éxitos en 10 pruebas, para una binomial con $p = 0,5$, es $0,04395 + 0,00977 + 0,00098 = 0,0547$. Por tanto, si los datos de cada uno de los 28 profesionales se analizan separadamente, contrastando $H_0: p = 0,5$ frente a $H_A: p > 0,5$, el menor de los 28 P valores sería 0,0547, lo que de nuevo no proporciona evidencia significativa a favor de H_A *.

Una forma diferente de realizar el análisis es investigar si las 280 observaciones que se presentan en la Tabla 13.2.4 son consistentes con un modelo binomial. En particular, podemos comprobar el modelo que establece que Y sí que es una distribución binomial, con $n = 10$ y $p = 0,5$, donde Y es el número de éxitos en 10 pruebas. (Esto es similar al análisis presentado en la Sección 3.9). En este caso se puede utilizar un test de bondad de ajuste. La Tabla 13.2.5 muestra los números observados (de la Tabla 13.2.4) y los números esperados para cada uno de los 11 posibles resultados. (Los números esperados no suman 28 debido al error de redondeo).

* Considerando el material de la Sección opcional 11.9 sobre comparaciones múltiples, nótese que si consideráramos todos los 28 tests, deberíamos requerir una buena dosis de evidencia antes de rechazar H_0 . Utilizando la corrección de Bonferroni, requeriríamos que el P valor fuera menor que $\alpha_{cw} = 0,05/28 = 0,0018$ para rechazar H_0 .

Tabla 13.2.5 Números observados y esperados (si $p = 0,5$) de éxitos en 10 pruebas en el experimento de toque terapéutico

Número de éxitos	Probabilidad binomial	Número observado, O	Número esperado, E
0	0,00098	0	0,027
1	0,00977	1	0,273
2	0,04395	1	1,231
3	0,11719	8	3,281
4	0,20508	5	5,742
5	0,24609	7	6,891
6	0,20508	2	5,742
7	0,11710	3	3,281
8	0,04395	1	1,231
9	0,00977	0	0,273
10	0,00098	0	0,027
Total	1,00000	28	27,999

El estadístico chi-cuadrado es $\chi_s^2 = \sum \frac{(o_i - e_i)^2}{e_i} = 11,7$. El estadístico de contraste tiene 10 grados de libertad, ya que hay 11 categorías en el modelo. El P valor para este test es 0,306, que es bastante grande. Por tanto, los datos son consistentes con una distribución binomial para la que $p = 0,5$ (es decir, los profesionales de TT podrían haber lanzado monedas para escoger una mano, en vez de intentar detectar el campo de energía humana del investigador). (Nota: estos datos no refutan la existencia del campo de energía humana; solo fallan en proporcionar evidencia a favor de su existencia). ■

Ejemplos breves

Consideraremos ahora algunos ejemplos en los que identificaremos el tipo de análisis que es apropiado, pero no realizaremos el análisis.

Ejemplo 13.2.6 Estrellas de mar

Unos investigadores midieron la longitud del brazo más largo en 200 miembros de la especie *Phataria unifascialis* (una estrella de mar que se encuentra en las aguas del Golfo de California, México). En una muestra de 184 individuos encontrados cerca de Loreto, la longitud media fue de 6,78 cm, con una DT de 1,21 cm. Para una muestra de 77 individuos encontrados cerca de Bahía de los Ángeles, la longitud media fue de 8,3 cm, con una DT de 1,33 cm⁶.

La variable de respuesta es numérica y hay dos grupos independientes. Por tanto, es apropiado un test t de dos muestras, junto con un intervalo de confianza para la diferencia de medias poblacionales. (Nota: la condición de normalidad no es esencial, ya que los tamaños muestrales son bastante grandes). ■

Ejemplo 13.2.7 Gemelos

Unos investigadores de Finlandia estudiaron los niveles de actividad física de cientos de parejas de gemelos del mismo sexo. En 1975 clasificaron los sujetos en las categorías de actividad física «deportista» y «sedentario». Realizaron un seguimiento de la salud de los sujetos hasta 1994, y en ese momento había varias parejas de gemelos en las que solo uno de ellos estaba vivo, ya que el otro había fallecido. En este grupo había 49 gemelos «sedentarios» que estaban vivos, pero cuya pareja «deportista» había fallecido. Había 76 gemelos «deportistas» que estaban vivos, pero cuya pareja «sedentario» había fallecido⁷.

La variable de respuesta en este estudio observacional es si un sujeto está o no vivo, una variable categórica. La variable de predicción es también categórica: si la persona es «sedentaria» o es «deportista». Como los datos están emparejados, es apropiado el test de McNemar. ■

Ejemplo 13.2.8 Muestras de suelo

Unos investigadores tomaron ocho muestras de suelo en seis ubicaciones de pastos mediterráneos. Dividieron las muestras en cuatro parejas y las pusieron en tiestos. Uno de los tiestos de cada pareja se regó constantemente, mientras que el otro se regó durante 13 días, después no se regó durante 18 días y finalmente se regó de nuevo durante 30 días. Los investigadores contabilizaron el número de germinaciones de cada tiesto durante el experimento⁸.

Este ejemplo es similar al Ejemplo 13.2.6, en el sentido de que hay dos muestras que se van a comparar y la variable de respuesta es numérica. Sin embargo, las muestras aquí están emparejadas, por tanto se requiere un análisis pareado (Capítulo 8). Si las 24 diferencias muestrales presentan una distribución normal, entonces se podría utilizar un test t o un intervalo de confianza para datos pareados. Si no, se podría intentar utilizar una transformación o emplear un test de signos o un test de rangos con signo de Wilcoxon. ■

Ejemplo 13.2.9 Vacunaciones

En 1996 hubo un brote de varicela en un centro de cuidado de niños de Georgia. Algunos de los niños habían sido vacunados contra la varicela, pero otros no. La varicela se presentó en 9 de los 66 niños vacunados y en 72 de los 82 niños no vacunados⁹.

Las variables de respuesta y de predicción en este caso son ambas categóricas. Los datos se podrían disponer en una tabla de contingencia de 2×2 y analizarlos con un test chi-cuadrado de independencia. La diferencia en las proporciones muestrales es obviamente bastante grande. Sin embargo, se trata de un estudio observacional y no de un experimento. Por tanto, no podemos concluir que la diferencia en las proporciones se debe enteramente al efecto de la vacuna, ya que los efectos de otras variables, como el estatus económico, se pueden confundir con el efecto de la vacuna. ■

Ejemplo 13.2.10 Estrógeno y esteroides

Se midieron los niveles en plasma de estrona más estradiol (Plasma E_{1+2}) en mujeres que recibían estrógeno (Premarin) y en un grupo de mujeres de control. Las mujeres que recibían estrógeno se dividieron en tres grupos de tratamiento. Un grupo recibió una dosis diaria de 0,625 mg, otro grupo recibió 1,25 mg y el tercer grupo recibió 2,5 mg. Los investigadores advirtieron que los niveles de esteroides en el plasma no estaban normalmente distribuidos, pero sí lo estaban después de aplicar una transformación logarítmica. Los datos se muestran, en escala logarítmica, en la Tabla 13.2.6¹⁰.

Tabla 13.2.6 Log ng/100 ml de concentración de E_{1+2} en el plasma para el estudio de estrógenos			
Grupo	n	Media	DT
Control	30	2,01	0,27
0,625	16	2,10	0,31
1,25	24	2,34	0,39
2,5	21	2,20	0,24

La variable de respuesta en este experimento, $\log(\text{concentración de } E_{1+2} \text{ en el plasma})$, es numérica. Ya ha sido transformada para que sea normal. Hay cuatro grupos independientes para ser comparados, de forma que es apropiado un análisis de la varianza. Un contraste que compare el control con el promedio de los tres grupos de tratamiento también sería útil. ■

Ejemplo 13.2.11 Caballitos del diablo

Un investigador capturó caballitos del diablo machos y los asignó aleatoriamente a tres grupos. En los del primer grupo los tamaños de las manchas rojas de las alas se aumentaron artificialmente con tinta roja. En los del segundo grupo las manchas de las alas se aumentaron con tinta clara. El tercer grupo sirvió como control. Después, se liberaron los caballitos del diablo en un área cerrada. Se determinó el número de supervivientes de cada uno de los tres grupos 22 días después. Había 312 caballitos del diablo en cada uno de los grupos. Después de 22 días, había 41 supervivientes en el grupo «aumento artificial con tinta roja», 49 supervivientes en el grupo «aumento con tinta clara» y 57 supervivientes en el grupo de control¹¹.

La variable de respuesta en este experimento, supervivencia, es categórica, así como la variable de predicción, estado/tipo de tinta. Los datos se podrían disponer en una tabla de contingencia de 2×3 y analizarlos con un test chi-cuadrado de independencia. ■

Ejemplo 13.2.12 Prevención del uso del tabaco

En el Proyecto Hutchinson para Prevención del Tabaco, 40 distritos escolares del estado de Washington se pusieron en 20 parejas basándose en su tamaño, su situación y la preponderancia del uso de tabaco en los institutos al comienzo de los estudios. En cada pareja, un distrito se asignó aleatoriamente a un grupo de intervención y el otro se asignó al grupo de control. Si un distrito escolar estaba en un grupo de intervención, entonces los estudiantes de tercer grado del distrito recibían un programa sobre la prevención del uso del tabaco y los profesores del distrito recibían formación especial para ayudar a los estudiantes a abstenerse de fumar. Esto se repitió un año más tarde con la siguiente promoción de estudiantes de tercer grado. Se realizó un seguimiento de todos los estudiantes durante varios años. Una medida principal del resultado del estudio fue si los estudiantes estaban o no fumando dos años después de graduarse en el instituto.

La unidad experimental aquí es un distrito escolar completo, por lo que es natural utilizar como variable de respuesta el porcentaje de estudiantes de un distrito que fuman, una variable numérica. La variable de predicción es categórica: grupo de intervención o grupo de control. Hay dos grupos, que están emparejados por el diseño del experimento. De las 20 parejas, hubo 13 en las que la proporción de fumadores era más alta en el distrito de control y 7 en las que la proporción de fumadores era más alta en el distrito de intervención¹². Se podría utilizar un test de signos para analizar estos datos. ■

Ejercicios 13.2.1-13.2.22

13.2.1 Unos investigadores realizaron un estudio clínico aleatorizado y doble ciego en el que algunos pacientes con esquizofrenia recibieron el medicamento clozapina y otros recibieron haloperidol. Tras un año, 61 de los 163 pacientes del grupo de la clozapina mostraron mejoras clínicamente importantes en los síntomas en comparación con 51 de 159 del grupo del haloperidol¹³. Identifique el tipo de método estadístico apropiado para estos datos, pero no realice realmente el análisis.

13.2.2 Considere los datos del Ejercicio 13.2.1. Realice un análisis apropiado completo de los datos que incluya una presentación gráfica y comentarios acerca de si los datos cumplen o no cumplen las condiciones de validez necesarias.

13.2.3 Un biólogo recogió datos sobre la altura (en pulgadas) y flujo en espiración de pico (FEP, una medida de cuánto aire puede espirar una persona, medido en l/min), para 10 mujeres¹⁴. Los datos fueron:

Sujeto	Altura	FEP	Sujeto	Altura	FEP
1	63	410	6	62	360
2	63	440	7	67	380
3	66	450	8	64	380
4	65	510	9	65	360
5	64	340	10	67	570

¿Está el FEP relacionado con la altura? Identifique el tipo de método estadístico apropiado para estos datos y esta pregunta, pero no realice realmente el análisis.

13.2.4 Considere los datos del Ejercicio 13.2.3. María es 1 pulgada más alta que Anika. Utilizando la información del Ejercicio 13.2.3, ¿cuánto se podría predecir que sería mayor el FEP de María con respecto al de Anika?

13.2.5 Un genetista autopolinizó plantas de boca de dragón de flores rosa y produjo una progenie de 97 plantas con los

siguientes colores: 22 plantas rojas, 52 plantas rosas y 23 plantas blancas¹⁵. El propósito del experimento era investigar un modelo genético que afirma que las probabilidades de rojo, rosa y blanco son 0,25, 0,50 y 0,25. Identifique el tipo de método estadístico apropiado para estos datos, pero no realice realmente el análisis.

13.2.6 Considere los datos del Ejercicio 13.2.5. Realice un análisis completo apropiado de los datos que incluya una presentación gráfica y comentarios acerca de si los datos cumplen o no cumplen las condiciones de validez necesarias.

13.2.7 El efecto de la dieta en las enfermedades de corazón ha sido ampliamente estudiado. Como parte de esta área general de investigación, unos investigadores estaban interesados en el efecto a corto plazo de la dieta sobre la función endotelial, como el efecto en el nivel de triglicéridos. Para estudiarlo, diseñaron un experimento en el que se suministró a 20 sujetos sanos, en orden aleatorio, un desayuno alto en grasas y un desayuno bajo en grasas a las 8 de la mañana, seguido de un ayuno de 12 horas, en días separados una semana entre sí. Se midieron los niveles de triglicéridos en suero de cada sujeto antes de cada desayuno y cuatro horas después de desayunar¹⁶. Si tuviera acceso a todas las medidas tomadas en el experimento, ¿cómo analizaría los datos?

13.2.8 Unos biólogos estaban interesados en la distribución de árboles en un área de bosque. Intentaron utilizar el número de árboles por parcela de 100 metros cuadrados como unidad de medida. Sin embargo, se dieron cuenta de que las formas de las parcelas podían afectar a la toma de datos. Para investigar la posibilidad, contaron el número de árboles en parcelas cuadradas, parcelas redondas y parcelas rectangulares. Los datos se muestran en la tabla siguiente¹⁷. ¿Qué tipo de análisis es apropiado para estos datos?

	Forma de la parcela		
	Cuadrada	Redonda	Rectangular
	5	5	10
	5	7	2
	5	5	3
	8	2	12
	8	4	9
	7	4	5
	4	4	3
	9	7	6
	9	7	5
	7	10	3
	5	9	8
	2	2	9
	8	7	3
Media	6,3	5,6	6,0
DT	2,14	2,47	3,27

13.2.9 Considere los datos del Ejercicio 13.2.8. Realice un análisis completo apropiado de los datos que incluya una presentación gráfica y comentarios acerca de si los datos cumplen o no cumplen las condiciones de validez necesarias.

13.2.10 Una muestra de 15 pacientes se dividió aleatoriamente en dos grupos como parte de un experimento doble ciego para comparar dos analgésicos¹⁸. Los 7 pacientes del primer grupo recibieron Demerol e informaron de los siguientes números de horas de alivio del dolor:

2, 6, 4, 13, 5, 8, 4

Los 8 pacientes del segundo grupo recibieron un medicamento experimental e informaron de los siguientes números de horas de alivio del dolor:

0, 8, 1, 4, 2, 2, 1, 3

¿Cómo se podrían analizar estos datos?

13.2.11 Considere los datos del Ejercicio 13.2.10. Realice un análisis completo apropiado de los datos que incluya una presentación gráfica y comentarios acerca de si los datos cumplen o no cumplen las condiciones de validez necesarias.

13.2.12 Un investigador estaba interesado en la relación entre la longitud del antebrazo y la altura. Midió las longitudes de los antebrazos y las alturas de una muestra de 16 mujeres y obtuvo los datos de la tabla siguiente¹⁹. ¿Cómo se podrían visualizar y analizar estos datos?

Antebrazo		Antebrazo	
Altura (cm)	Longitud (cm)	Altura (cm)	Longitud (cm)
163	25,5	157	26
161	26	178	27
151	25	163	24,5
163	25	161	26
166	27,2	173	28
168	26	160	24,5
170	26	158	25
163	26	170	26

13.2.13 Se realizó un estudio clínico aleatorizado y doble ciego sobre pacientes que tenían angioplastia coronaria para comparar el medicamento lovastatin con un placebo. Se midió el porcentaje de estenosis (estrechamiento de los vasos sanguíneos) después de la angioplastia en 160 pacientes que recibieron lovastatin y en 161 pacientes que recibieron un placebo. En el grupo del lovastatin el promedio fue del 46 %, con una DT del 20 %. En el grupo del placebo el promedio fue de 44 %, con una DT del 21 %²⁰. ¿Qué tipo de análisis es apropiado para estos datos?

13.2.14 Considere los datos del Ejercicio 13.2.13.

- Realice un análisis apropiado de los datos.
- Describa un procedimiento gráfico para visualizar estos datos.
- Comente cómo sería probable que los datos cumplieran las condiciones necesarias para la validez incluso aunque no tenga acceso a los datos iniciales.

13.2.15 Unos investigadores estudiaron a personas que habían recibido inmunoglobulina intravenosa (IGIV) para ver si habían desarrollado infecciones del virus de la hepatitis C (VHC). En parte de sus análisis, consideraron dosis de Gammagard (un producto de IGIV) que fueron recibidas por 210 pacientes. Dividieron los pacientes en cuatro grupos de acuerdo al número de dosis de «Gammagard proveniente de plasma no examinado o examinado anti-VHC de primera generación». Entre 48 personas que recibieron de 0 a 3 dosis, hubo 4 casos de infección por VHC. Hubo 2 casos de infección por VHC entre 45 personas que recibieron de 4 a 20 dosis, hubo 7 casos de infección por VHC entre 57 personas que recibieron entre 21 y 65 dosis y hubo 10 casos de infección por VHC entre las 51 personas que recibieron más de 65 dosis²¹. ¿Qué tipo de análisis es adecuado para estos datos?

13.2.16 Considere los datos del Ejercicio 13.2.15. Realice un análisis apropiado de los datos.

13.2.17 Se realizó un experimento para estudiar el efecto del tamoxifeno en pacientes con cáncer del cuello del útero. Una de las medidas realizadas, tanto antes como después de suministrar el tamoxifeno, fue la densidad de microvasos (DMV). DMV, que mide el número de vasos por milímetro cuadrado, es una medida que se relaciona con la formación de vasos sanguíneos que alimentan un tumor y le permiten crecer y extenderse. Por tanto, valores pequeños de DMV son mejores que valores grandes. Se muestran a continuación los datos de 18 pacientes²². ¿Cómo se podrían analizar estos datos?

Paciente	DMV antes	DMV después	Paciente	DMV antes	DMV después
1	98	75	10	70	60
2	100	60	11	60	65
3	82	25	12	88	45
4	100	55	13	45	36
5	93	78	14	159	144
6	119	102	15	65	27
7	70	58	16	98	90
8	78	70	17	66	16
9	104	90	18	67	53

13.2.18 Considere los datos del Ejercicio 13.2.17. Realice un análisis completo apropiado de los datos que incluya una presentación gráfica y comentarios acerca de si los datos cumplen o no cumplen las condiciones de validez necesarias.

13.2.19 Como parte de un gran experimento, unos investigadores plantaron 2.400 liquidámbares, 2.400 sicomoros y 1.200 fresnos verdes. Tras 18 años, las tasas de supervivencia fueron del 93 % para los liquidámbares, del 88 % para los sicomoros y del 95 % para los fresnos verdes²³. ¿Qué tipo de análisis es apropiado para estos datos?

13.2.20 Considere los datos del Ejercicio 13.2.19. Realice un análisis completo apropiado de los datos que incluya una presentación gráfica y comentarios acerca de si los datos cumplen o no cumplen las condiciones de validez necesarias.

13.2.21 Un grupo de estudiantes universitarios se dividieron en tres grupos de acuerdo con su fuerza corporal de levantamiento. La fuerza de sus piernas se probó midiendo cuántas veces podían levantar unas pesas de pierna de 246 libras antes del agotamiento. (Se permitía solo un segundo de descanso entre levantamientos consecutivos). Los datos se muestran en la tabla siguiente²⁴. ¿Qué tipo de análisis es apropiado para estos datos?

	Fuerza corporal de levantamiento		
	Baja	Media	Alta
	55	40	181
	70	200	85
	45	250	416
	246	192	228
	240	117	257
	96	215	316
	225		134
Media	140	169	231
TD	93	77	112

13.2.22 Considere los datos del Ejercicio 13.2.21. Realice un análisis completo apropiado de los datos que incluya una presentación gráfica y comentarios acerca de si los datos cumplen o no cumplen las condiciones de validez necesarias.

Notas

1. Datos no publicados cortesía de M. A. Johnson y F. Bretos. Datos recogidos en Oberlin College en el otoño de 1997.
2. Fish, F. E. (1998). Comparative kinematics and hydrodynamics of odontocete cetaceans: morphological and ecological correlates with swimming performance. *Journal of Experimental Biology* **201**, 2867-2877. Los datos presentados aquí están tomados de la Figura 1 del artículo.
3. Fisher, B., Costantino, J. P., Wickerman, L., Redmond, C. K., Kavanah, M., Cronin, W. M., Vogel, V., Robidoux, A., Dimitrov, N., Atkins, J., Daly, M., Wieand, S., Tan-Chiu, E., Ford, L., Wolmark, N., et al. (1998). Tamoxifen for prevention of breast cancer: Report of the national surgical adjuvant breast and bowel project P-1 study. *Journal of the National Cancer Institute* **30**, 1371-1388.
4. Datos no publicados cortesía de K. Pretl. Datos recogidos en Oberlin College en la primavera de 1997.
5. Rosa, L., Rosa, E., Sarner, L. y Barrett, S. (1998). A close look at therapeutic touch. *Journal of the American Medical Association* **279**, 1005-1010.
6. Morgan, M. B. y Cowles, D. L. (1996). The effects of temperature on the behaviour and physiology of *Phataria unifascialis* (Gray) (Echinodermata, Asteroidea); implications for the species' distribution in the Gulf of California, Mexico. *Journal of Experimental Marine Biology and Ecology* **208**, 13-27.
7. Kujala, U. M., Kaprio, J., Sarna, S. y Koskenvuo, M. (1998). Relationship of leisure-time physical activity and mortality: The Finnish cohort study. *Journal of the American Medical Association* **279**, 440-444. La categoría «deportista» tenía realmente dos subcategorías de «deportista ocasional» y «deportista de preparación», que hemos combinado aquí.
8. Espigares, T. y Peco, B. (1995). Mediterranean annual pasture dynamics: Impact of autumn drought. *Journal of Ecology* **83**, 135-142.
9. Izurieta, H. S., Strebel, P. M. y Blake, P. A. (1997). Postlicensure effectiveness of varicella vaccine during an outbreak in a child care center. *Journal of the American Medical Association* **278**, 1495-1499.
10. Rose, D. P., Fern, M., Liskowski, L. y Milbrath, J. R. (1977). Effect of treatment with estrogen conjugates on endogenous plasma steroids. *Obstetrics and Gynecology* **49**, 80-82.
11. Grether, G. (1997). Survival cost of an intrasexually selected ornament in a damselfly. *Proceedings of the Royal Society of London* **B 264**, 207-210.
12. Peterson, A. V., Kealey, K. A., Mann, S. L., Marek, P. M. y Sarason, I. G. (2000). Hutchinson smoking prevention project: Long-term randomized trial in school-based tobacco use prevention-results on smoking. *Journal of the National Cancer Institute* **92**, 1979-1991. Los autores no analizaron estos datos con un test de signos, sino con un test de aleatorización, similar al presentado en la Sección 7.1.
13. Rosenheck, R., Cramer, J., Xu, W., Thomas, J., Henderson, W., Frisman, L., Fye, C. y Charney, D., for the Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia (1997). A comparison of clozapine and haloperidol in hospitalized patients with refractory schizophrenia. *New England Journal of Medicine* **337**, 809-815.
14. Datos no publicados cortesía de K. Roberts. Datos recogidos en Oberlin College en la primavera de 1997.
15. Baur, E., Fischer, E. y Lenz, F. (1931). *Human Heredity*, 3.^a ed. New York: Macmillan, p. 45.
16. Plotnick, G. D., Corretti, M. C. y Vogel, R. A. (1997). Effect of antioxidant vitamins on the transient impairment of endothelium-dependent brachial artery vasoactivity following a single high-fat meal. *Journal of the American Medical Association* **278**, 1682-1686.
17. Datos no publicados tomados por E. Lohan y M. Josephy, Oberlin College, en el otoño de 1997.
18. Meier, P., Free, S. M. y Jackson, G. L. (1958). Reconsideration of methodology in studies of pain relief. *Biometrics* **14**, 330-342.
19. Datos no publicados recogidos por J. Amundson, Oberlin College, en el otoño de 1997.
20. Weintraub, W. S., Boccuzzi, S. J., Klein, J. L., Kosinski, A. S., King, S. B., Ivanhoe, R., Cedarholm, J. C., Stillabower, M. E., Talley, J. D., DeMaio, S. J., O'Neill, W. W., Frazier, J. E., Cohen-Bernstein, C. L., Robbins, D. C., Brown, C. L., Alexander, R. W. y the Lovastatin Restenosis Trial Study Group. (1994). Lack of effect of lovastatin on restenosis after coronary angioplasty. *New England Journal of Medicine* **331**, 1331-1337.
21. Bresee, J., Mast, E. E., Coleman, P. J., Baron, M. J., Schonberger, L. B., Alter, M. J., Jonas, M. M., Yu, M. W., Renzi, P. M. y Schneider, L. C. (1996). Hepatitis C virus infection associated with administration of intravenous immune globulin: A cohort study. *Journal of the American Medical Association* **276**, 1563-1567.
22. Ferrandina, G., Ranelletti, F. O., Larocca, L. M., Maggiano, N., Fruscella, E., Legge, F., Santeusano, G., Bombonati, A., Mancuso, S. y Scambia, G. (2001). Tamoxifen modulates the expression of Ki67, apoptosis, and microvessel density in cervical cancer. *Cancer Clinical Research* **7**, 2656-2661.
23. Devine, W. D., Houston, A. E. y Tyler, D. D. (2000). Growth of three hardwood species through 18 years on a former agricultural bottom-land. *Southern Journal of Applied Forestry* **24**, 159-165.
24. Datos no publicados recogidos por S. Haaz, Oberlin College, en la primavera de 1999. La división en grupos se basa en la fuerza de levantamiento, ajustada con respecto a la altura, al peso, y a si el estudiante estaba o no en un equipo de atletismo.

APÉNDICES

Apéndice

- 3.1 Más sobre la fórmula de la distribución binomial 573
- 3.2 Media y desviación típica de la distribución binomial 575
- 4.1 Áreas de regiones de extensión infinita 576
- 5.1 Relación entre el Teorema Central del Límite y la aproximación normal a la distribución binomial 577
- 6.1 Dígitos significativos 578
- 7.1 Cálculo de la potencia 579
- 7.2 Más sobre el test de Wilcoxon-Mann-Whitney 580
- 9.1 Más sobre intervalos de confianza para proporciones 582
- 12.1 Fórmulas de mínimos cuadrados 583
- 12.2 Obtención de la Propiedad 12.3.1 585

Apéndice 3.1

Más sobre la fórmula de la distribución binomial

En este apéndice explicamos más sobre el razonamiento que hay tras la fórmula de la distribución binomial.

La fórmula de la distribución binomial

Comenzaremos obteniendo la fórmula de la distribución binomial para $n = 3$. Supongamos que realizamos tres pruebas independientes y que cada prueba puede tener como resultado: éxito (E) o fracaso (F). Supongamos además que en cada prueba las probabilidades de éxito y fracaso son

$$\Pr\{S\} = p$$

$$\Pr\{F\} = 1 - p$$

Hay ocho posibles resultados de las tres pruebas. Razonando como en el Ejemplo 3.6.3, se puede ver que las probabilidades de esos resultados son las siguientes:

Resultado	Número de éxitos	Número de fracasos	Probabilidad
FFF	0	3	$(1 - p)^3$
FFE	1	2	$p(1 - p)^2$
FEF	1	2	$p(1 - p)^2$
EFF	1	2	$p(1 - p)^2$
FEE	2	1	$p^2(1 - p)$
EFE	2	1	$p^2(1 - p)$
EEF	2	1	$p^2(1 - p)$
EEE	3	0	p^3

Razonando de nuevo de forma paralela al Ejemplo 3.6.3, esas probabilidades se pueden combinar para obtener la fórmula de la distribución binomial para $n = 3$ como se muestra en la tabla:

Número de		Probabilidad
Éxitos, j	Fracasos, $n - j$	
0	3	$1p^0(1 - p)^3$
1	2	$3p^1(1 - p)^2$
2	1	$3p^2(1 - p)^1$
3	0	$1p^3(1 - p)^0$

Esta distribución ilustra el origen de los coeficientes binomiales. El coeficiente ${}_3C_1 (= 3)$ es el número de formas en las que se pueden disponer 2 E y 1 F. El coeficiente ${}_3C_2 (= 3)$ es el número de formas en las que se pueden disponer 1 E y 2 F.

Un argumento similar a este permite ver que la fórmula general (para cualquier n) es

$$\Pr\{j \text{ éxitos y } n - j \text{ fracasos}\} = {}_nC_j p^j (1 - p)^{n-j}$$

con

$${}_nC_j = \text{número de formas en las que se pueden disponer } j \text{ E y } (n - j) \text{ F}$$

Combinaciones El coeficiente binomial ${}_nC_j$ se puede interpretar también como el número de combinaciones de n elementos tomados de j en j . Es igual al número de subconjuntos diferentes de tamaño j que se pueden formar en un conjunto con n elementos.

Coeficientes binomiales: una fórmula

Los coeficientes binomiales se pueden calcular mediante la fórmula

$${}_nC_j = \frac{n!}{j!(n - j)!}$$

donde $x!$ (« x factorial») se define para cualquier entero positivo x como

$$x! = x(x - 1)(x - 2) \cdots (2)(1)$$

y $0! = 1$.

Por ejemplo, para $n = 7$ y $j = 4$, la fórmula da como resultado

$$\begin{aligned} {}_7C_4 &= \frac{7!}{4!3!} = \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(4 \times 3 \times 2 \times 1)(3 \times 2 \times 1)} \\ &= 35 \end{aligned}$$

Para ver por qué esto es correcto, consideremos en detalle por qué el número de formas de disponer 4 E y 3 F debe ser igual a

$$\frac{7!}{4!3!}$$

Supongamos que los 4 E y los 3 F se escriben en tarjetas, como estas:

$$\boxed{E_1} \quad \boxed{E_2} \quad \boxed{E_3} \quad \boxed{E_4} \quad \boxed{F_1} \quad \boxed{F_2} \quad \boxed{F_3}$$

Pondremos temporalmente subíndices a los E y a los F para distinguirlos. Veamos primero cuántas formas hay de disponer 7 cartas en fila:

Hay 7 opciones para la carta que va primera;
 para cada una de estas, hay 6 opciones para la carta que va segunda;
 para cada una de estas, hay 5 opciones para la carta que va tercera;
 para cada una de estas, hay 4 opciones para la carta que va cuarta;
 para cada una de estas, hay 3 opciones para la carta que va quinta;
 para cada una de estas, hay 2 opciones para la carta que va sexta;
 para cada una de estas, hay 1 opción para la carta que va última;

Se deduce por tanto que hay $7!$ formas de disponer las 7 cartas. Consideremos ahora las posiciones de los 4 E. Hay $4!$ formas en las que se pueden ordenar los E entre ellos. Asimismo, hay $3!$ formas en las que se pueden disponer los F entre ellos. Si ignoráramos los subíndices de los E y los F, entonces algunas de las $7!$ formas de ordenar las cartas serían indistinguibles. De hecho, cualquier reordenación de los E *entre ellos mismos* no afecta al aspecto de la disposición de las 7 cartas. De forma similar, cualquier reordenación de los F *entre ellos mismos* no afecta al aspecto de la disposición de las 7 cartas. Por tanto, el número de disposiciones *distinguibles* de las 7 cartas es

$$\frac{7!}{4!3!}$$

Apéndice 3.2

Media y desviación típica de la distribución binomial

Supongamos que Y es una variable aleatoria binomial con n pruebas y que p es la probabilidad de éxito en cada prueba. Podemos ver Y como la suma de n variables $X_1, X_2, \dots, X_n$, donde cada X_i es 0 o 1 (0 para un fracaso y 1 para un éxito). Es decir, $Y = \sum X_i$, con $\Pr\{X_i = 0\} = 1 - p$ y $\Pr\{X_i = 1\} = p$. Las n muestras X_i son una muestra aleatoria de una población hipotética de X cuya media es $\mu_X = p$ (ya que $0 \times (1 - p) + 1 \times p = p$).

Consideremos ahora la desviación típica poblacional, σ_X , de la población de X . Recordemos, de la Sección 2.8, que dada una variable X , la definición de σ es

$$\sigma = \sqrt{\text{Valor poblacional medio de } (X - \mu)^2}$$

Para la población de X , la media es $\mu_X = p$. Por tanto, para esta población,

$$\sigma_X = \sqrt{\text{Valor poblacional medio de } (X - p)^2}$$

En la población de X , $(X - p)^2$ toma solo dos posibles valores

$$(X - p)^2 = \begin{cases} (0 - p)^2 & \text{si } X = 0 \\ (1 - p)^2 & \text{si } X = 1 \end{cases}$$

Además, estos valores aparecen en las proporciones $(1 - p)$ y p , respectivamente, por lo que el valor medio poblacional de $(X - p)^2$ es igual a

$$(0 - p)^2 \times (1 - p) + (1 - p)^2 \times p$$

Esto se puede simplificar:

$$\begin{aligned}
 p^2 \times (1 - p) + (1 - p)^2 \times p &= p\{p(1 - p) + (1 - p)^2\} \\
 &= p\{p - p^2 + 1 - 2p + p^2\} \\
 &= p(1 - p)
 \end{aligned}$$

Por consiguiente, el valor de la media poblacional de $(X - p)^2$ es $p(1 - p)$, por lo que $\sigma_X = \sqrt{p(1 - p)}$.

La variable aleatoria binomial Y es $\sum X_i$. Para obtener la media y la desviación típica de Y , es necesario conocer dos propiedades:

Propiedad 1: Para cualquier conjunto de variables aleatorias $X_1, X_2, \dots, X_n$, la media de $\sum X_i = \sum (\text{media de } X_i)$.

Propiedad 2: Dado un conjunto de variables aleatorias independientes $X_1, X_2, \dots, X_n$, la varianza de $\sum X_i = \sum (\text{varianza de } X_i)$.

(Recuérdese que la varianza, σ^2 , es el cuadrado de la desviación típica, σ).

Utilizando la Propiedad 1, podemos ver que la media de Y es la media de $\sum X_i$, que es $\sum p$. Por tanto, la media de Y es $\mu_Y = np$.

Utilizando la Propiedad 2, la varianza de Y es la varianza de $\sum X_i$, que es igual a $\sum (\text{varianza de } X_i)$ o $np(1 - p)$. Por tanto, la desviación típica de Y es $\sigma_Y = \sqrt{np(1 - p)}$.

Apéndice 4.1

Áreas de regiones de extensión infinita

Consideremos una región acotada entre una curva normal y el eje horizontal. Como la curva nunca toca el eje, la región se extiende hasta el infinito por la derecha y por la izquierda. Pero sin embargo el área de la región es igual a 1,0. ¿Cómo es posible que una región que se extiende hasta el infinito tenga un área finita?

Para comprender esta situación paradójica, consideremos la Figura A.4.1, que muestra una región más simple que la acotada por una curva normal. En esta región, la anchura de cada barra es 1,0, la altura de la primera barra es $\frac{1}{2}$, la altura de la segunda barra es la mitad de la barra de la primera, la de la tercera es la mitad de la barra de la segunda, y así sucesivamente. Las barras forman una región que se extiende hasta el infinito. No obstante, veremos que tiene sentido decir que el área de la región completa es igual a 1,0.

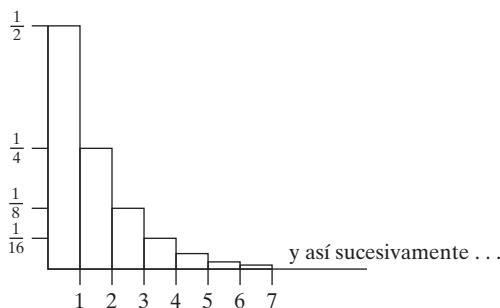


Figura A.4.1

Consideremos primero el área de cada una de las barras. El área de la primera barra es $\frac{1}{2}$, el área de la segunda barra es $\frac{1}{4}$, el área de la tercera es $\frac{1}{8}$, y así sucesivamente. Supongamos ahora que escogemos un número, por ejemplo k , y sumamos las áreas de las primeras k barras, como se indica en la tabla de la página siguiente.

El área total de las primeras dos barras es $\frac{3}{4}$, la de las primeras tres barras es $\frac{7}{8}$, y así sucesivamente. De hecho, el área total de las primeras k barras es

$$\frac{2^k - 1}{2^k} = 1 - \frac{1}{2^k}$$

Si k es muy grande, el área será muy cercana a 1,0. De hecho, podemos aproximar el área a 1,0 tanto como queramos, simplemente eligiendo un valor de k suficientemente grande. En esas circunstancias, es razonable decir que el área total de la región completa, de extensión infinita, es exactamente igual a 1,0.

El ejemplo anterior muestra que el área de una región de extensión infinita puede ser finita. Asimismo, el área total bajo la curva normal es 1,0 (pero la demostración de este hecho requiere cálculo bastante avanzado).

Barra	Altura de la barra	Área total acumulada
1	$\frac{1}{2}$	$\frac{1}{2}$
2	$\frac{1}{4}$	$\frac{3}{4}$
3	$\frac{1}{8}$	$\frac{7}{8}$
4	$\frac{1}{16}$	$\frac{15}{16}$
$\vdots$	$\vdots$	$\vdots$
k	$\frac{1}{2^k}$	$\frac{2^k - 1}{2^k}$

Apéndice 5.1

Relación entre el Teorema Central del Límite y la aproximación normal a la distribución binomial

Consideremos el muestreo de una población dicotómica. El Teorema 5.4.1 establece que la distribución muestral de $\hat{P}$, y la distribución binomial equivalente, se pueden aproximar por distribuciones normales. En este apéndice mostraremos cómo se relacionan estas aproximaciones con el Teorema 5.2.1 y con el Teorema Central del Límite.

Como se muestra en el Apéndice 3.2, si Y es una variable aleatoria binomial con n pruebas con una probabilidad p de éxito en cada prueba, entonces podemos ver Y como una suma de n variables $X_1, X_2, \dots, X_n$, donde cada X_i es igual a 0 o a 1 (0 para un fallo y 1 para un éxito). Dada una población de 0 y 1 donde la proporción de 1 está dada por p , la media es p y la desviación típica es $\sigma = \sqrt{p(1-p)}$. La media muestral de $X_1, X_2, \dots, X_n$ es $\bar{X}$ que es igual a la proporción de 1 en la muestra (es decir, $\hat{P}$). Por tanto, la proporción muestral $\hat{P}$ se puede ver como una media muestral y, por tanto, su distribución muestral queda descrita por el Teorema 5.2.1.

Del apartado 3 del Teorema 5.2.1 (el Teorema Central del Límite), se deduce que la distribución muestral de $\hat{P}$ es aproximadamente normal si n es grande. Del apartado 1 del Teorema 5.2.1, se deduce que la media de la distribución muestral de $\hat{P}$ es igual a la media poblacional (es decir, p). Este valor está dado por el Teorema 5.4.1(b). Del apartado 2 del Teorema 5.2.1, se deduce que la desviación típica de la distribución muestral de $\hat{P}$ es igual a

$$\frac{\sigma}{\sqrt{n}}$$

donde σ representa la desviación típica de la población de 0 y 1, que es $\sqrt{p(1-p)}$. Por tanto, la desviación típica de la distribución muestral de $\hat{P}$ es igual a

$$\frac{\sqrt{p(1-p)}}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$$

que es el valor dado en el Teorema 5.4.1(b).

Es preciso señalar que la distribución binomial es una versión restablecida de la distribución de muestreo de $\hat{P}$: $\hat{P} = \frac{Y}{n}$, de manera que $Y = n\hat{P}$. Por tanto, la distribución binomial también se puede aproximar mediante una curva normal con una media restablecida y una desviación típica adecuadas. La media de $\hat{P}$ es p y la DT de $\hat{P}$ es

$$\sqrt{\frac{p(1-p)}{n}}$$

La media reescalada es np y la desviación típica reescalada es

$$n \sqrt{\frac{p(1-p)}{n}} = \sqrt{np(1-p)}$$

que son las dadas en el Teorema 5.4.1(a).

Apéndice 6.1

Dígitos significativos

En este apéndice revisamos el concepto de dígitos significativos. Comencemos con un ejemplo.

Supongamos que el rector de una universidad informa de que en su universidad hay 23.000 estudiantes. ¿Cuántos dígitos significativos hay en el número

23.000?

Cuando el número se expresa de esta forma (en notación ordinaria en vez de científica), no es realmente posible decir con seguridad cuántos dígitos significativos tiene. ¿Quiso decir *realmente* el rector

23.000 y no 23.001 ni 22.999?

Si es así, entonces los cinco dígitos son significativos. Si (como es probable) quiso decir

23.000 y no 22.000 ni 24.000

entonces solo el 2 y el 3 son dígitos significativos, ya que solo esos dígitos se conocen con certeza. Los tres 0 en 23.000 son marcadores de sustitución. La notación científica resuelve esta ambigüedad:

$2,3 \times 10^4$ tiene 2 dígitos significativos

$2,3000 \times 10^4$ tiene 5 dígitos significativos

Como ilustra el ejemplo anterior, se puede clarificar el número de dígitos significativos de un número expresándolo en notación científica. He aquí algunos ejemplos:

Notación ordinaria	Notación científica	Número de dígitos significativos
60.700	$6,07 \times 10^4$	3
60.700	$6,0700 \times 10^4$	5
60,7	$6,07 \times 10^1$	3
60,70	$6,070 \times 10^1$	4
0,0607	$6,07 \times 10^{-2}$	3
0,06070	$6,070 \times 10^{-2}$	4

En los números anteriores, nótese que el cero del interior (entre 6 y 7) es siempre un dígito significativo. Los ceros a la izquierda del 6 no son significativos. Los ceros del final (después del 7) son significativos en notación científica y ambiguos en notación ordinaria. Los dígitos distintos de cero son siempre significativos.

He aquí algunos ejemplos de redondeo de un número a dos dígitos significativos.

Número	Redondeo a dos dígitos significativos
60.700	61.000 (es decir, $6,1 \times 10^4$)
60,7	61
0,0607	0,061
0,0592	0,059
0,0596	0,060

Apéndice 7.1

Cálculo de la potencia

Los tamaños muestrales requeridos que se dan en la Tabla 5 se determinan calculando la potencia del test t . Para muestras grandes, un cálculo adecuado de la potencia se puede basar en la curva normal (Tabla 3). En este apéndice indicamos cómo se realiza este cálculo aproximado.

Recuérdese que la potencia es la probabilidad de rechazar H_0 cuando H_A es cierta. Por tanto, para calcular la potencia, es necesario conocer la distribución muestral de t_s cuando H_A es cierta. Para muestras grandes, la distribución muestral se puede aproximar mediante una curva normal, como indica el siguiente teorema.

Teorema A.1 Supongamos que tomamos muestras aleatorias independientes, cada una de ellas de tamaño n , de poblaciones normales de medias μ_1 y μ_2 y desviación típica común σ . Si n es grande, la distribución muestral de t_s se puede aproximar mediante una distribución normal con

$$\text{Media} = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}} = \sqrt{\frac{n}{2}} \left(\frac{\mu_1 - \mu_2}{\sigma} \right)$$

y

$$\text{Desviación típica} = 1$$

Para ilustrar el uso del Teorema A.1 en el cálculo de potencias, supongamos que estamos considerando un test t de una cola con $\alpha = 0,025$. Las hipótesis son

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 > \mu_2$$

Si deseamos una potencia de 0,80 para un tamaño del efecto de 0,4, entonces la Tabla 5 recomienda muestras de tamaño $n = 100$. Confirmemos esta recomendación utilizando el Teorema A.1.

Si H_0 es cierta, de forma que $\mu_1 = \mu_2$, entonces la distribución muestral de t_s es aproximadamente normal con media igual a 0 y DT igual a 1. Esta es la distribución nula de t_s , que se muestra como la curva discontinua en la Figura A.7.1.

Supongamos que de hecho H_A es cierta, que el tamaño del efecto es

$$\frac{\mu_1 - \mu_2}{\sigma} = 0,4$$

y que estamos utilizando muestras de tamaño $n = 100$. Entonces, de acuerdo con el Teorema A.1, la distribución muestral de t_s será aproximadamente normal con DT igual a 1 y media igual a

$$\sqrt{\frac{n}{2}} \left(\frac{\mu_1 - \mu_2}{\sigma} \right) = \sqrt{\frac{100}{2}} (0,4) = 2,83$$

Esta distribución es la curva continua de la figura.

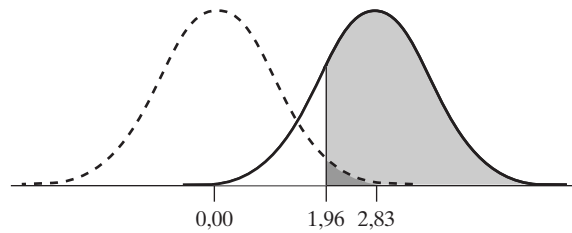


Figura A.7.1

Para $n_1 = n_2 = 100$, tenemos $gl \approx \infty$, por lo que vemos en la Tabla 4 que el valor crítico es igual a 1,96. Por tanto, el P valor sería menor que 0,025 y rechazaríamos H_0 si

$$t_s > 1,96$$

Utilizando la curva discontinua, la probabilidad de este suceso es igual a 0,025, que se muestra en la figura como el área en gris oscuro. Utilizando la curva continua, la probabilidad de que $t_s > 1,96$ incluye toda el área sombreada de la figura. Esa área sombreada se puede determinar a partir de la Tabla 3 utilizando

$$Z = \frac{1,96 - 2,38}{1} = -0,87$$

De la Tabla 3, se obtiene que el área es $0,8078 \approx 0,81$. Por tanto, hemos demostrado que, para $n_1 = n_2 = 100$,

$$\text{si } \frac{\mu_1 - \mu_2}{\sigma} = 0,4, \text{ entonces } \Pr\{\text{rechazar } H_0\} \approx 0,81$$

Hemos obtenido que la potencia contra la alternativa especificada es aproximadamente igual a 0,81. Esto está bastante de acuerdo con lo que indica la Tabla 5, que dice que la potencia es igual a 0,80.

Si estuviéramos realizando un test de dos colas con $\alpha = 0,05$, el valor crítico sería de nuevo 1,96 y, por tanto, la potencia sería de nuevo aproximadamente igual a 0,81, ya que el área bajo la curva continua correspondiente a la cola izquierda de la curva discontinua es despreciable.

Por supuesto, al construir la Tabla 5, uno empieza con la potencia especificada (0,80) y determina n , en vez de hacerlo de la forma anterior. Este problema «inverso» se puede resolver utilizando un planteamiento similar al precedente. En la figura, el área sombreada (0,80) es el dato. Esto determina el valor de Z y a su vez determina n , una vez se ha especificado el tamaño del efecto.

Apéndice 7.2

Más sobre el test de Wilcoxon-Mann-Whitney

En la Sección 7.10 vimos cómo se relacionan los valores críticos del test de Wilcoxon-Mann-Whitney con la distribución nula de K_1 , K_2 y U_s . En este apéndice indicamos cómo se pueden determinar dichas distribuciones nulas por métodos simples de conteo.

Consideremos los tamaños muestrales de $n = 5$ y $n' = 4$. En la Figura 7.10.3, los Y_1 e Y_2 se muestran como puntos. Para ahorrar espacio, representaremos los datos de forma más compacta: representaremos cada Y_1 como un «1» y cada Y_2 como un «2». Por tanto, la disposición de la Figura 7.10.3(a) (donde los Y_1 están todos a la izquierda de los Y_2) se representaría como

1 1 1 1 1 2 2 2 2

Para tamaños muestrales $n = 5$ y $n' = 4$, hay 126 posibles disposiciones de los Y_1 e Y_2 . He aquí una lista parcial de esas disposiciones y los valores asociados de K_1 y K_2 :

Número	Disposición	K_1	K_2
1	1 1 1 1 1 2 2 2 2	0	20
2	1 1 1 1 2 1 2 2 2	1	19
3	1 1 1 1 2 2 1 2 2	2	18
4	1 1 1 2 1 1 2 2 2	2	18
5	1 1 2 1 1 1 2 2 2	3	17
6	1 1 1 2 1 2 1 2 2	3	17
7	1 1 1 1 2 2 2 1 2	3	17
8	1 2 1 1 1 1 2 2 2	4	16
9	1 1 2 1 1 2 1 2 2	4	16
10	1 1 1 2 1 2 2 1 2	4	16
11	1 1 1 2 2 1 1 2 2	4	16
12	1 1 1 1 2 2 2 2 1	4	16
	... y así sucesivamente ...		
126	2 2 2 2 1 1 1 1 1	20	0

Para determinar las distribuciones nulas de esta lista, necesitamos saber la probabilidad de las diferentes disposiciones, asumiendo que H_0 es cierta. De acuerdo con H_0 , las 9 observaciones (Y) se toman aleatoriamente de la misma población. Bajo este supuesto, se puede demostrar que las 126 disposiciones son *igualmente probables*. Debido a este simple y elegante hecho, la distribución nula de K_1 y K_2 (y, por lo tanto, de U_s) se pueden determinar por conteo directo. Trabajando con la lista anterior, tenemos las siguientes probabilidades:

K_1	K_2	Probabilidad
0	20	$\frac{1}{126}$
1	19	$\frac{1}{126}$
2	18	$\frac{2}{126}$
3	17	$\frac{3}{126}$
4	16	$\frac{5}{126}$
	... y así sucesivamente ...	
20	0	$\frac{1}{126}$
		Total 1

Estas probabilidades forman la distribución nula de K_1 y K_2 [representadas en la Figura 7.10.4(a)]. Por ejemplo, la primera probabilidad de la distribución nula es

$$\Pr\{K_1 = 0, K_2 = 20\} = \frac{1}{126} \approx 0,008$$

como se indica en la Sección 7.10.

¿Por qué es el test de Wilcoxon-Mann-Whitney independiente de la distribución? La razón debería quedar clara después de la exposición anterior. Si las dos distribuciones poblacionales son las mismas, entonces todas las posibles disposiciones de las Y son igualmente probables y la forma específica de la distribución poblacional no importa (excepto, por supuesto en que hemos supuesto que no hay empates, ya que las distribuciones nulas se alteran si hay empates).

La distribución nula de Wilcoxon-Mann-Whitney siempre se puede determinar de forma directa mediante el método de conteo comentado anteriormente (aunque para tamaños muestrales grandes el conteo es muy tedioso y se usan en su lugar métodos aproximados). El número de posibles disposiciones para muestras de tamaño n y n' es igual a

$$\frac{(n + n')!}{n!n'!}$$

Por ejemplo, para tamaños muestrales de 5 y 4 (como antes), obtenemos

$$\frac{9!}{5!4!}$$

(Para ver por qué esta fórmula funciona, remitimos al lector a la exposición de la fórmula de los coeficientes binomiales al final del Apéndice 3.1; el razonamiento es completamente paralelo).

Apéndice 9.1

Más sobre intervalos de confianza para proporciones

En este apéndice presentamos algunos detalles técnicos sobre el intervalo de confianza para proporciones presentado en la Sección 9.2. Para una exposición más completa de estas ideas, véase el artículo de Agresti y Coull que se cita en la Nota 3 del Capítulo 9.

Supongamos que deseamos obtener un intervalo de confianza $100(1 - \alpha)\%$ para una proporción p . Por ejemplo, un nivel de confianza del 95 % corresponde a $\alpha = 0,05$. Un procedimiento común es utilizar

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

que se denomina *intervalo de confianza de Wald*.

Otra forma de construir un intervalo de confianza es obtener todos los valores de p para los que

$$-z_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}} \leq z_{\alpha/2} \quad (1)$$

(Esto se conoce como «inversión del contraste de hipótesis». El contraste de hipótesis para proporciones se explica en la Sección 9.4). La idea básica aquí es que la distribución muestral de $\hat{p}$ se puede aproximar por una distribución normal y que tomamos los valores correspondientes a la zona central de probabilidad $100(1 - \alpha)\%$ (es decir, el 95 % central) de la distribución normal.

En la presentación que sigue z indicará $z_{\alpha/2}$. Despejando p en la desigualdad (1) se obtiene un intervalo de la forma

$$\frac{\hat{p} + \frac{z^2}{2n} \pm z \sqrt{\frac{\hat{p}(1 - \hat{p}) + \frac{z^2}{4n}}{n}}}{1 + \frac{z^2}{n}}$$

que se denomina *intervalo de confianza de puntuación*.

La mayoría de los libros presentan el intervalo de confianza de Wald (sin darle ese nombre), ya que tiene una forma mucho más simple que el de puntuación. Sin embargo, el intervalo de confianza de Wald tiene pobres propiedades de

cobertura: un intervalo de Wald con una confianza nominal del 95 % realmente podría contener a p solo el 80 % de las veces, en vez del 95 %. El intervalo de confianza de puntuación tiene excelentes propiedades de cobertura, pero es bastante complejo.

La formulación del *intervalo de confianza de Wilson* presentada en la Sección 9.2 se basa en aproximar el intervalo de puntuación. Realizando operaciones, se puede demostrar que el punto medio del intervalo de puntuación es

$$\hat{p}\left(\frac{n}{n+z^2}\right) + \frac{1}{2}\left(\frac{z^2}{n+z^2}\right) \quad (2)$$

que es la media ponderada de $\hat{p}$ y $\frac{1}{2}$, con pesos $\frac{n}{n+z^2}$ y $\frac{z^2}{n+z^2}$. Nótese que cuando n aumenta, se da más peso a $\hat{p}$. Para n pequeño, se da más peso a $\frac{1}{2}$. Como $\hat{p} = \frac{y}{n}$, el punto medio dado por la fórmula (2) es

$$\frac{y}{n+z^2} + \frac{\frac{1}{2}z^2}{n+z^2}$$

Para un intervalo de confianza del 95 %, $z = 1.96 \approx 2$, por lo que el punto medio se aproxima como

$$\frac{y}{n+4} + \frac{2}{n+4} = \frac{y+2}{n+4}$$

que en la Sección 9.1 denominamos $\tilde{p}$.

El error típico que se utiliza en el intervalo de confianza de Wald es $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Esto se basa en el hecho de que la varianza de Y es igual a $np(1-p)$, como se obtuvo en el Apéndice 3.2, de modo que la varianza de $\hat{P}$ es igual a $\frac{np(1-p)}{n^2}$ o $\frac{p(1-p)}{n}$. Por tanto, la desviación típica de $\hat{P}$ es $\sqrt{\frac{p(1-p)}{n}}$, como se explicó en el Apéndice 5.1. El error típico usa la proporción muestral $\hat{p}$ en vez del valor desconocido de p . Asimismo, la varianza de $\tilde{P}$ es igual a $\frac{np(1-p)}{(n+z^2)^2}$, que es aproximadamente $\frac{p(1-p)}{n+z^2}$. Por tanto, el error típico del intervalo de confianza de Wilson está dado por $\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+z^2}}$. Para un intervalo de confianza del 95 %, $z \approx 2$, con lo que

$$\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

Como nota final, mencionaremos que algunos autores defienden el uso de un intervalo de confianza «exacto», basado en la distribución binomial. Sin embargo, como demuestran Agresti y Coull, el intervalo de confianza exacto es bastante conservador, lo que es una razón para que no sea ampliamente utilizado (es también bastante complejo).

Apéndice 12.1

Fórmulas de mínimos cuadrados

En este apéndice demostramos que el criterio de mínimos cuadrados conduce a las fórmulas

$$b_1 = r \frac{s_y}{s_x} \quad \text{y} \quad b_0 = \bar{y} - b_1 \bar{x}$$

Haremos uso del hecho de que el mínimo de una función $Q(x) = Ax^2 + Bx + C$ cuadrática está en $x = -\frac{B}{2A}$.

Resultado preliminar

Dado un conjunto de datos $y_1, y_2, \dots, y_n$, el número c que minimiza $\sum_{i=1}^n (y_i - c)^2$ es la media $\bar{y}$.

Para verlo, desarrollamos primero $(y_i - c)^2$ para obtener $y_i^2 - 2y_i c + c^2$ y después distribuimos el sumatorio, obteniendo

$$\sum_{i=1}^n (y_i - c)^2 = \sum_{i=1}^n y_i^2 - 2c \sum_{i=1}^n y_i + nc^2$$

Esta última expresión es una función cuadrática $Q(c) = Ac^2 + Bc + C$, siendo $A = n$ y $B = -2 \sum_{i=1}^n y_i$. Por tanto, el mínimo se produce en $c = -\frac{B}{2A} = -\frac{-2 \sum_{i=1}^n y_i}{2n} = \bar{y}$.

Recuérdese que si $\hat{y}_i = b_0 + b_1 x_i$ es la recta de regresión de mínimos cuadrados, entonces b_0 y b_1 son los valores que minimizan la suma de cuadrados residual, dada por

$$SC(\text{resid}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2$$

Podemos expresar $y_i - (b_0 + b_1 x_i)$ como $(y_i - b_1 x_i) - b_0$. Seguidamente, aplicamos el resultado preliminar con $(y_i - b_1 x_i)$ en el lugar de y_i y b_0 en el lugar de c . Por tanto, el mínimo de la suma de cuadrados residual se produce cuando

$$b_0 = \text{la media de } (y_i - b_1 x_i) = \bar{y} - b_1 \bar{x} \quad (1)$$

Por tanto, la recta de mínimos cuadrados pasa por el punto de los promedios $(\bar{x}, \bar{y})$.

Sustituyendo el valor de b_0 en la suma de cuadrados residual se obtiene

$$\begin{aligned} SC(\text{resid}) &= \sum_{i=1}^n (y_i - (\bar{y} - b_1 \bar{x}) - b_1 x_i)^2 = \sum_{i=1}^n (y_i - \bar{y} + b_1 \bar{x} - b_1 x_i)^2 \\ &= \sum_{i=1}^n [(y_i - \bar{y}) - b_1(x_i - \bar{x})]^2 \end{aligned}$$

que se puede desarrollar como

$$\begin{aligned} SC(\text{resid}) &= \sum_{i=1}^n [(y_i - \bar{y})^2 - 2b_1(x_i - \bar{x})(y_i - \bar{y}) + b_1^2(x_i - \bar{x})^2] \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2b_1 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Por tanto, la suma de cuadrados residual es una función cuadrática de b_1 con $A = \sum_{i=1}^n (x_i - \bar{x})^2$ y $B = -2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$. Por tanto, el mínimo de $SC(\text{resid})$ ocurre en $x = -\frac{B}{2A}$ o

$$\begin{aligned} b_1 &= -\frac{-2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{2 \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{(n-1)s_x^2} s_x s_y \\ &= r \left(\frac{s_y}{s_x} \right) \end{aligned} \quad (2)$$

Las ecuaciones (1) y (2) en conjunto proporcionan las fórmulas de los coeficientes de la recta de regresión de mínimos cuadrados.

Apéndice 12.2

Obtención de la Propiedad 12.3.1

Por definición

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

Definamos $SPXY = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, $SCX = \sum_{i=1}^n (x_i - \bar{x})^2$, y $SCY = \sum_{i=1}^n (y_i - \bar{y})^2$. Entonces

$$s_x^2 = \frac{SCX}{n-1} \quad y \quad s_y^2 = \frac{SCY}{n-1}$$

$$\text{Entonces } r = \frac{1}{n-1} \frac{SPXY}{s_x s_y} \quad y \quad r^2 = \frac{SPXY^2}{(SCX)(SCY)}.$$

Recuérdese que $b_1 = r \left(\frac{s_y}{s_x} \right)$ y $b_0 = \bar{y} - b_1 \bar{x}$. Por tanto,

$$b_1 = \frac{1}{n-1} \frac{SPXY}{s_x s_y} \left(\frac{s_y}{s_x} \right) = \frac{1}{n-1} \frac{SPXY}{s_x^2} = \frac{SPXY}{SCX}$$

La definición de residuo es $y - \hat{y}$, pero esto es lo mismo que $y - (b_0 + b_1 x) = y - [(\bar{y} - b_1 \bar{x}) + b_1 x] = (y - \bar{y}) - b_1(x - \bar{x})$.

Por tanto,

$$\begin{aligned} SC(\text{resid}) &= \sum_{i=1}^n [(y_i - \bar{y}) - b_1(x_i - \bar{x})]^2 \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - 2b_1 \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) + b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= SCY - 2b_1 SPXY + b_1^2 SCX \\ &= SCY - 2 \frac{SPXY}{SCX} SPXY + \frac{SPXY^2}{SCX} \\ &= SCY - \frac{SPXY^2}{SCX} \end{aligned}$$

Por consiguiente, hemos establecido que $\frac{SPXY^2}{SCX} = SCY - SC(\text{resid})$.

Entonces

$$r^2 = \frac{SPXY^2}{(SCX)(SCY)} = \frac{SCY - SC(\text{resid})}{SCY}$$

Así,

$$r^2 = \frac{\frac{SCY}{n-1} - \frac{SC(\text{resid})}{n-1}}{\frac{SCY}{n-1}} = \frac{s_y^2 - s_e^2 \left(\frac{n-2}{n-1} \right)}{s_y^2} = \frac{s_y^2 - s_e^2 f}{s_y^2} \approx \frac{s_y^2 - s_e^2}{s_y^2}$$

siendo $f = \frac{n-2}{n-1}$.

El factor f es cercano a 1 a menos que n sea muy pequeño. He aquí algunos valores de f :

n	f
10	0,89
15	0,93
20	0,95

La aproximación es razonablemente buena si $n \geq 10$.

TABLAS ESTADÍSTICAS

Tabla 1	Dígitos aleatorios	587	Tabla 7	Valores críticos y P valores de B_s para el test de signos	601
Tabla 2	Coeficientes binomiales ${}_nC_j$	591	Tabla 8	Valores críticos y P valores de W_s para el test de rangos con signo de Wilcoxon	602
Tabla 3	Áreas bajo la curva normal	592	Tabla 9	Valores críticos de la distribución chi-cuadrado	603
Tabla 4	Valores críticos de la distribución t de Student	594	Tabla 10	Valores críticos de la distribución F	604
Tabla 5	Tamaños muestrales necesarios para niveles de potencia seleccionados para tests t de muestras independientes	595	Tabla 11	Multiplicadores de Bonferroni para intervalos de confianza del 95 %	614
Tabla 6	Valores críticos y P valores de U_s para el test de Wilcoxon-Mann-Whitney	597			

TABLA 1 Dígitos aleatorios

	01	06	11	16	21	26	31	36	41	46
01	06048	96063	22049	86532	75170	65711	29969	06826	39208	80631
02	25636	73908	85512	78073	19089	66458	06597	93985	14193	69366
03	61378	45410	43511	54364	97334	01267	28304	35047	38789	84896
04	15919	71559	12310	00727	54473	51547	09816	83641	72973	75367
05	47328	20405	88019	82276	33679	10328	25116	59176	64675	95141
06	72548	80667	53893	64400	81955	15163	06146	58549	75530	19582
07	87154	04130	55985	44508	37515	71689	80765	46598	45539	12792
08	68379	96636	32154	94718	22845	80265	92747	66238	58474	23783
09	89391	54041	70806	36012	30833	83132	39338	54753	00722	44568
10	15816	60231	28365	61924	66934	21243	09896	92428	51611	46756
11	29618	55219	18394	11625	27673	08117	89314	42581	36897	03738
12	30723	42988	30002	95364	45473	46107	34222	00739	84847	49096
13	54028	04975	92323	53836	76128	84762	32050	59516	40831	59687
14	40376	02036	48087	05216	26684	97959	85601	86622	70750	15603
15	64439	37357	90935	57330	79738	65361	85944	23619	30504	61564
16	83037	30144	29166	20915	53462	42573	75204	50064	08847	07082
17	71071	01636	31085	71638	77357	14256	89174	15184	81701	21592
18	67891	43187	58159	24144	29683	04276	02987	04571	18334	04291
19	52487	39499	97330	40045	47304	98528	00422	82693	87547	73525
20	67550	82107	27302	79145	73213	27217	19211	59784	63929	04609
21	86472	80165	70773	90519	49710	31921	36102	45042	04203	01439
22	08699	38051	60404	06609	98435	91560	22634	98014	43316	61099
23	59596	13000	07655	74837	81211	71530	28341	83110	72289	25180
24	31810	54868	92799	09893	97499	96509	71548	06462	40498	22628
25	71753	90756	21382	84209	95900	11119	34507	61241	17641	83147

(Continúa)

TABLA 1 Dígitos aleatorios (*continuación*)

	51	56	61	66	71	76	81	86	91	96
01	64825	74126	86159	26710	49256	04655	06001	73192	67463	16746
02	46184	63916	89160	87844	53352	43318	70766	23625	09906	65847
03	79976	48891	69431	86571	25979	58755	08884	36704	01107	12308
04	10656	47210	48512	06805	42114	98741	51440	06070	49071	02700
05	18058	84528	56753	02623	81077	60045	06678	53748	10386	37895
06	58979	98046	88467	27762	24781	12559	98384	40926	79570	34746
07	12705	41974	14473	49872	29368	80556	95833	20766	76643	35656
08	39660	83664	18592	82388	27899	24223	36462	61582	95173	36155
09	00360	42077	84161	04464	45042	29560	37916	29889	00342	82533
10	09873	64084	34685	53542	09254	23257	14713	44295	94139	00403
11	12957	84063	79808	23633	77133	41422	26559	29131	74402	82213
12	06090	71584	48965	60201	02786	88929	19861	99361	27535	38297
13	66812	57167	28185	19708	74672	25615	61640	18955	40854	50749
14	91701	36216	66249	04256	31694	33127	67529	73254	72065	74294
15	02775	78899	36471	37098	50270	58933	91765	95157	01384	75388
16	75892	53340	92363	58300	77300	08059	63743	12159	05640	87014
17	18581	70057	82031	68349	55759	46851	33632	28855	74633	08598
18	69698	18177	52824	61742	58119	04168	57843	37870	50988	80316
19	30023	30731	00803	09336	87709	39307	09732	66031	04904	91929
20	94334	05698	97910	37850	77074	56152	67521	48973	29448	84115
21	64133	14640	28418	45405	86974	06666	07879	54026	92264	23418
22	93895	83557	17326	28030	09113	56793	79703	18804	75807	20144
23	54438	83097	52533	86245	02182	11746	58164	90520	99255	44830
24	90565	76710	42456	22612	00232	18919	24019	32254	30703	00678
25	90848	81871	24382	16218	98216	42323	75061	68261	09071	68776

TABLA 1 Dígitos aleatorios (*continuación*)

	01	06	11	16	21	26	31	36	41	46
26	17155	07370	65655	04824	53417	20737	70510	92615	89967	50216
27	36211	24724	94769	16940	43138	25260	75318	69037	95982	28631
28	94777	66946	16120	56382	58416	92391	81457	28101	69766	32436
29	52994	58881	81841	51844	75566	48567	18552	66829	91230	39141
30	84643	32635	51440	96854	35739	66440	82806	82841	56302	31640
31	95690	34873	11297	60518	72717	47616	55751	37187	31413	31132
32	64093	92948	21565	51686	40368	66151	82877	99951	85069	54503
33	89484	50055	67586	16439	96385	67868	66597	51433	44764	66573
34	70184	38164	74646	90244	83169	85276	07598	69242	90088	32308
35	75601	91867	80848	94484	98532	36183	28549	17704	28653	80027
36	99044	78699	34681	31049	40790	50445	79897	68203	11486	93676
37	10272	18347	89369	02355	76671	34097	03791	93817	43142	24974
38	69738	85488	34453	80876	43018	59967	84458	71906	54019	70023
39	93441	58902	17871	45425	29066	04553	42644	54624	34498	27319
40	25814	74497	75642	58350	64118	87400	82870	26143	46624	21404
41	29757	84506	48617	48844	35139	97855	43435	74581	35678	69793
42	56666	86113	06805	09470	07992	54079	00517	19313	53741	25306
43	26401	71007	12500	27815	86490	01370	47826	36009	10447	25953
44	40747	59584	83453	30875	39509	82829	42878	13844	84131	48524
45	99434	51563	73915	03867	24785	19324	21254	11641	25940	92026
46	50734	88330	39128	14261	00584	94266	99677	19852	49673	18680
47	89728	32743	19102	83279	68308	41160	32365	25774	39699	50743
48	71395	61945	41082	93648	99874	82577	26507	07054	29381	16995
49	50945	68182	23108	95765	81136	06792	13322	41631	37118	35881
50	36525	26551	28457	75699	74537	68623	50099	91909	23508	35751

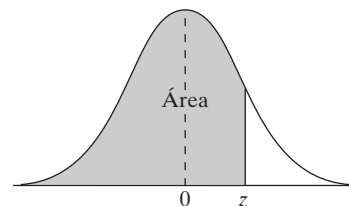
TABLA 1 Dígitos aleatorios (*continuación*)

	51	56	61	66	71	76	81	86	91	96
26	41169	08175	69938	61958	72578	31791	74952	71055	40369	00429
27	84627	70347	41566	00019	24481	15677	54506	54545	89563	50049
28	67460	49111	54004	61428	61034	47197	90084	88113	39145	94757
29	99231	60774	52238	05102	71690	72215	61323	13326	01674	81510
30	95775	73679	04900	27666	18424	59793	14965	22220	30682	35488
31	42179	98675	69593	17901	48741	59902	98034	12976	60921	73047
32	91196	05878	92346	45886	31080	21714	19168	94070	77375	10444
33	18794	03741	17612	65467	27698	20456	91737	36008	88225	58013
34	88311	93622	34501	70402	12272	65995	66086	04938	52966	71909
35	17904	33710	42812	72105	91848	39724	26361	09634	50552	98769
36	05905	28509	69631	69177	39081	58818	01998	53949	47884	91326
37	23432	22211	65648	71866	49532	45529	00189	80025	68956	26445
38	29684	43229	54771	90604	48938	13663	24736	83199	41512	43364
39	26506	65067	64252	49765	87650	72082	48997	04845	00136	98941
40	08807	43756	01579	34508	94082	68736	67149	00209	76138	95467
41	50636	70304	73556	32872	07809	20787	85921	41748	10553	97988
42	32437	41588	46991	36667	98127	05072	63700	51803	77262	31970
43	32571	97567	78420	04633	96574	88830	01314	04811	10904	85923
44	28773	22496	11743	23294	78070	20910	86722	50551	37356	92698
45	65768	76188	07781	05314	26017	07741	22268	31374	53559	46971
46	68601	06488	73776	45361	89059	59775	59149	64095	10352	11107
47	98364	17663	85972	72263	93178	04284	79236	04567	31813	82283
48	95308	70577	96712	85697	55685	19023	98112	96915	50791	31107
49	68681	24419	15362	60771	09962	45891	03130	09937	15775	51935
50	30721	22371	65174	57363	37851	71554	19708	23880	86638	05880

TABLA 2 Coeficientes binomiales ${}_nC_j$

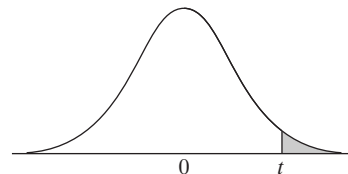
n	j										
	0	1	2	3	4	5	6	7	8	9	10
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1.287	1.716	1.716	1.287	715	286
14	1	14	91	364	1.001	2.002	3.003	3.432	3.003	2.002	1.001
15	1	15	105	455	1.365	3.003	5.005	6.435	6.435	5.005	3.003
16	1	16	120	560	1.820	4.368	8.008	11.440	12.870	11.440	8.008
17	1	17	136	680	2.380	6.188	12.376	19.448	24.310	24.310	19.448
18	1	18	153	816	3.060	8.568	18.564	31.824	43.758	48.620	43.758
19	1	19	171	969	3.876	11.628	27.132	50.388	75.582	92.378	92.378
20	1	20	190	1.140	4.845	15.504	38.760	77.520	125.970	167.960	184.756

TABLA 3 Áreas bajo la curva normal



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,9	0,0019	0,0018	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641

TABLA 4 Valores críticos de la distribución *t* de Student



gl	Probabilidad de la cola superior									
	0,20	0,10	0,05	0,04	0,03	0,025	0,02	0,01	0,005	0,0005
1	1,376	3,078	6,314	7,916	10,579	12,706	15,895	31,821	63,657	636,619
2	1,061	1,886	2,920	3,320	3,896	4,303	4,849	6,965	9,925	31,599
3	0,978	1,638	2,353	2,605	2,951	3,182	3,482	4,541	5,841	12,924
4	0,941	1,533	2,132	2,333	2,601	2,776	2,999	3,747	4,604	8,610
5	0,920	1,476	2,015	2,191	2,422	2,571	2,757	3,365	4,032	6,869
6	0,906	1,440	1,943	2,104	2,313	2,447	2,612	3,143	3,707	5,959
7	0,896	1,415	1,895	2,046	2,241	2,365	2,517	2,998	3,499	5,408
8	0,889	1,397	1,860	2,004	2,189	2,306	2,449	2,896	3,355	5,041
9	0,883	1,383	1,833	1,973	2,150	2,262	2,398	2,821	3,250	4,781
10	0,879	1,372	1,812	1,948	2,120	2,228	2,359	2,764	3,169	4,587
11	0,876	1,363	1,796	1,928	2,096	2,201	2,328	2,718	3,106	4,437
12	0,873	1,356	1,782	1,912	2,076	2,179	2,303	2,681	3,055	4,318
13	0,870	1,350	1,771	1,899	2,060	2,160	2,282	2,650	3,012	4,221
14	0,868	1,345	1,761	1,888	2,046	2,145	2,264	2,624	2,977	4,140
15	0,866	1,341	1,753	1,878	2,034	2,131	2,249	2,602	2,947	4,073
16	0,865	1,337	1,746	1,869	2,024	2,120	2,235	2,583	2,921	4,015
17	0,863	1,333	1,740	1,862	2,015	2,110	2,224	2,567	2,898	3,965
18	0,862	1,330	1,734	1,855	2,007	2,101	2,214	2,552	2,878	3,922
19	0,861	1,328	1,729	1,850	2,000	2,093	2,205	2,539	2,861	3,883
20	0,860	1,325	1,725	1,844	1,994	2,086	2,197	2,528	2,845	3,850
21	0,859	1,323	1,721	1,840	1,988	2,080	2,189	2,518	2,831	3,819
22	0,858	1,321	1,717	1,835	1,983	2,074	2,183	2,508	2,819	3,792
23	0,858	1,319	1,714	1,832	1,978	2,069	2,177	2,500	2,807	3,768
24	0,857	1,318	1,711	1,828	1,974	2,064	2,172	2,492	2,797	3,745
25	0,856	1,316	1,708	1,825	1,970	2,060	2,167	2,485	2,787	3,725
26	0,856	1,315	1,706	1,822	1,967	2,056	2,162	2,479	2,779	3,707
27	0,855	1,314	1,703	1,819	1,963	2,052	2,158	2,473	2,771	3,690
28	0,855	1,313	1,701	1,817	1,960	2,048	2,154	2,467	2,763	3,674
29	0,854	1,311	1,699	1,814	1,957	2,045	2,150	2,462	2,756	3,659
30	0,854	1,310	1,697	1,812	1,955	2,042	2,147	2,457	2,750	3,646
40	0,851	1,303	1,684	1,796	1,936	2,021	2,123	2,423	2,704	3,551
50	0,849	1,299	1,676	1,787	1,924	2,009	2,109	2,403	2,678	3,496
60	0,848	1,296	1,671	1,781	1,917	2,000	2,099	2,390	2,660	3,460
70	0,847	1,294	1,667	1,776	1,912	1,994	2,093	2,381	2,648	3,435
80	0,846	1,292	1,664	1,773	1,908	1,990	2,088	2,374	2,639	3,416
100	0,845	1,290	1,660	1,769	1,902	1,984	2,081	2,364	2,626	3,390
140	0,844	1,288	1,656	1,763	1,896	1,977	2,073	2,353	2,611	3,361
1.000	0,842	1,282	1,646	1,752	1,883	1,962	2,056	2,330	2,581	3,300
∞	0,842	1,282	1,645	1,751	1,881	1,960	2,054	2,326	2,576	3,291
	60 %	80 %	90 %	92 %	94 %	95 %	96 %	98 %	99 %	99,9 %
	Valor crítico para nivel de confianza del									

TABLA 5 Tamaños muestrales necesarios para niveles de potencia seleccionados para tests t de muestras independientes

Potencia→	Nivel de significación (test de dos colas)																					
	$\alpha = 0,01$					$\alpha = 0,02$					$\alpha = 0,05$					$\alpha = 0,10$						
	0,99	0,95	0,90	0,80	0,50	0,99	0,95	0,90	0,80	0,50	0,99	0,95	0,90	0,80	0,50	0,99	0,95	0,90	0,80	0,50		
0,20															124					137	0,20	
0,25																				88	0,25	
0,30										123						87				61	0,30	
0,35				110						90						64			102	45	0,35	
0,40				85						70					100			108	78	35	0,40	
0,45			118	68						101					105			108	86	62	28	0,45
0,50			96	55						82					86			88	70	51	23	0,50
$\frac{ \mu_1 - \mu_2 }{\sigma}$										106					106							
										88					71							
										68					53							
										38					27							
	0,55									106					87							
										74					60							
	0,60									32					45							
										58					23							
										27					20							
	0,65									49					39							
									34					30								
									29					17								
0,70									43					34								
									24					17								
									21					15								
0,75									38					23								
									26					11								
									19					10								
0,80									33					21								
									19					14								
									26					12								
0,85									30					18								
									17					9								
									23					8								
0,90									15					16								
									11					8								
									21					15								
									10					7								
0,95									10					15								
									10					7								
									9					7								
1,00									9					7								
									14					14								
									17					14								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
									13					9								
	</																					

(Continúa)

TABLA 5 Tamaños muestrales necesarios para niveles de potencia seleccionados para tests *t* de muestras independientes
(*continuación*)

Potencia→	Nivel de significación (test de dos colas)																								
	$\alpha = 0,01$					$\alpha = 0,02$										$\alpha = 0,05$					$\alpha = 0,10$				
	0,99	0,95	0,90	0,80	0,50	0,99	0,95	0,90	0,80	0,50	0,99	0,95	0,90	0,80	0,50	0,99	0,95	0,90	0,80	0,50					
1,1	42	32	27	22	13	38	28	23	19	11	32	23	19	14	8	27	19	15	12	6					
1,2	36	27	23	18	11	32	24	20	16	9	27	20	16	12	7	23	16	13	10	5					
1,3	31	23	20	16	10	28	21	17	14	8	23	17	14	11	6	20	14	11	9	5					
1,4	27	20	17	14	9	24	18	15	12	8	20	15	12	10	6	17	12	10	8	4					
1,5	24	18	15	13	8	21	16	14	11	7	18	13	11	9	5	15	11	9	7	4					
1,6	21	16	14	11	7	19	14	12	10	6	16	12	10	8	5	14	10	8	6	4					
1,7	19	15	13	10	7	17	13	11	9	6	14	11	9	7	4	12	9	7	6	3					
1,8	17	13	11	10	6	15	12	10	8	5	13	10	8	6	4	11	8	7	5						
1,9	16	12	11	9	6	14	11	9	8	5	12	9	7	6	4	10	7	6	5						
2,0	14	11	10	8	6	13	10	9	7	5	11	8	7	6	4	9	7	6	4						
σ	2,1	13	10	9	8	5	12	9	8	7	5	10	8	6	5	3	8	6	5	4					
	2,2	12	10	8	7	5	11	9	7	6	4	9	7	6	5		8	6	5	4					
	2,3	11	9	8	7	5	10	8	7	6	4	9	7	6	5		7	5	5	4					
	2,4	11	9	8	6	5	10	8	7	6	4	8	6	5	4		7	5	4	4					
	2,5	10	8	7	6	4	9	7	6	5	4	8	6	5	4		6	5	4	3					
3,0	8	6	6	5	4	7	6	5	4	3	6	5	4	4		5	4	3							
3,5	6	5	5	4	3	6	5	4	4		5	4	4	3		4	3								
4,0	6	5	4	4		5	4	4	3		4	4	3			4									
	$\alpha = 0,005$					$\alpha = 0,01$					$\alpha = 0,025$					$\alpha = 0,05$									

Fuente: «Number of observations for t-test of difference between two means.» *Research*, Volume 1 (1948), pp. 520-525. Utilizado con permiso de the Longman Group UK Ltd. y Butterworth Scientific Publications.

TABLA 6 Valores críticos y P valores de U_s para el test de Wilcoxon-Mann-Whitney

Nota: debido a que la distribución nula del test de Wilcoxon-Mann-Whitney es discreta, esta tabla proporciona valores seleccionados del estadístico de contraste U_s **en negrilla** y los correspondientes P valores para una alternativa no direccional *en cursiva*. Los P valores direccionales se obtienen dividiendo por dos los números en cursiva.

n	n'	0,20	0,10	0,05	0,025	0,01	0,005
3	2	6 0,200					
	3	8 0,200	9 0,100				
4	2	8 0,133					
	3	11 0,114	12 0,057				
5	4	13 0,200	15 0,057	16 0,029			
	2	9 0,191	10 0,095				
6	3	13 0,143	14 0,071	15 0,036			
	4	16 0,191	18 0,064	19 0,032	20 0,016		
7	5	20 0,151	21 0,095	23 0,032	24 0,016	25 0,0079	
	2	11 0,143	12 0,071				
8	3	15 0,167	16 0,095	17 0,048	18 0,024		
	4	19 0,171	21 0,067	22 0,038	23 0,019	24 0,0095	
9	5	23 0,178	25 0,082	27 0,030	28 0,017	29 0,0087	30 0,0043
	6	27 0,180	29 0,093	31 0,041	33 0,015	34 0,0087	35 0,0043
10	2	13 0,111	14 0,056				
	3	17 0,183	19 0,067	20 0,033	21 0,017		
11	4	22 0,164	24 0,072	25 0,042	26 0,024	28 0,0061	
	5	27 0,149	29 0,073	30 0,048	32 0,018	34 0,0051	35 0,0025
12	6	31 0,181	34 0,073	36 0,035	37 0,022	39 0,0082	40 0,0047
	7	36 0,165	38 0,097	41 0,038	43 0,018	45 0,0070	46 0,0041
13	2	14 0,178	15 0,089	16 0,044			
	3	19 0,194	21 0,085	22 0,049	23 0,024		
14	4	25 0,154	27 0,073	28 0,049	30 0,016	31 0,0081	32 0,0040
	5	30 0,171	32 0,093	34 0,045	36 0,019	38 0,0062	39 0,0031
15	6	35 0,181	38 0,081	40 0,043	42 0,020	44 0,0080	45 0,0047
	7	40 0,189	43 0,094	46 0,041	48 0,021	50 0,0093	52 0,0037
16	8	45 0,195	49 0,083	51 0,050	54 0,021	57 0,0070	58 0,0047
	2	16 0,146	17 0,073	18 0,036			
17	3	22 0,146	23 0,100	25 0,036	26 0,018	27 0,0091	
	4	27 0,199	30 0,076	32 0,034	33 0,020	35 0,0056	36 0,0028
18	5	33 0,190	36 0,083	38 0,042	40 0,019	42 0,0070	43 0,0040
	6	39 0,181	42 0,088	44 0,050	47 0,018	49 0,0076	50 0,0048
19	7	45 0,174	48 0,091	51 0,042	53 0,023	56 0,0079	58 0,0033
	8	50 0,200	54 0,093	57 0,046	60 0,021	63 0,0079	65 0,0037
20	9	56 0,190	60 0,094	64 0,040	66 0,024	70 0,0078	72 0,0040
	2	17 0,182	19 0,061	20 0,030			
21	3	24 0,161	26 0,077	27 0,049	29 0,014	30 0,0070	
	4	30 0,188	33 0,076	35 0,036	36 0,024	38 0,0080	39 0,0040
22	5	37 0,165	39 0,099	42 0,040	44 0,019	46 0,0080	47 0,0047
	6	43 0,181	46 0,093	49 0,042	51 0,023	54 0,0075	55 0,0047
23	7	49 0,193	53 0,088	56 0,043	58 0,025	61 0,0097	63 0,0046
	8	56 0,173	60 0,083	63 0,043	66 0,021	69 0,0085	71 0,0044
24	9	62 0,182	66 0,095	70 0,044	73 0,022	77 0,0076	79 0,0041
	10	68 0,190	73 0,089	77 0,043	80 0,023	84 0,0089	87 0,0039

(Continúa)

TABLA 6 Valores críticos y *P* valores de U_s para el test de Wilcoxon-Mann-Whitney (continuación)

<i>n</i>	<i>n'</i>	0,20	0,10	0,05	0,025	0,01	0,005
11	2	19 0,154	21 0,051	22 0,026			
	3	26 0,170	28 0,088	30 0,039	31 0,022	33 0,0055	
	4	33 0,177	36 0,078	38 0,040	40 0,018	42 0,0059	43 0,0029
	5	40 0,180	43 0,090	46 0,038	48 0,019	50 0,0087	52 0,0032
	6	47 0,180	50 0,098	53 0,048	56 0,020	59 0,0071	60 0,0048
	7	54 0,179	58 0,085	61 0,044	64 0,020	67 0,0083	69 0,0041
	8	61 0,177	65 0,091	69 0,041	72 0,020	75 0,0091	77 0,0050
	9	68 0,175	72 0,095	76 0,047	80 0,020	83 0,0097	86 0,0042
	10	74 0,197	79 0,099	84 0,043	87 0,024	92 0,0079	94 0,0048
	11	81 0,193	87 0,088	91 0,047	95 0,023	100 0,0083	103 0,0041
12	2	20 0,198	22 0,088	23 0,044	24 0,022		
	3	28 0,180	31 0,070	32 0,048	34 0,018	35 0,0088	36 0,0044
	4	36 0,170	39 0,078	41 0,042	43 0,020	45 0,0077	46 0,0044
	5	43 0,195	47 0,082	49 0,049	52 0,019	54 0,0094	56 0,0039
	6	51 0,180	55 0,083	58 0,042	60 0,025	63 0,0097	65 0,0047
	7	58 0,196	63 0,083	66 0,045	69 0,022	72 0,0098	75 0,0037
	8	66 0,181	70 0,098	74 0,047	78 0,020	81 0,0096	84 0,0041
	9	73 0,193	78 0,096	82 0,049	86 0,023	90 0,0093	93 0,0043
	10	81 0,180	86 0,093	91 0,043	94 0,025	99 0,0090	102 0,0044
	11	88 0,190	94 0,091	99 0,044	103 0,023	108 0,0086	111 0,0045
	12	95 0,198	102 0,089	107 0,045	111 0,024	117 0,0083	120 0,0045
13	2	22 0,171	24 0,076	25 0,038	26 0,019		
	3	30 0,189	33 0,082	35 0,039	36 0,025	38 0,0071	39 0,0036
	4	39 0,163	42 0,079	44 0,045	46 0,023	49 0,0059	50 0,0034
	5	47 0,173	50 0,095	53 0,046	56 0,019	58 0,0098	60 0,0044
	6	55 0,179	59 0,087	62 0,046	65 0,022	68 0,0092	70 0,0047
	7	63 0,183	67 0,097	71 0,046	74 0,024	78 0,0085	80 0,0047
	8	71 0,185	76 0,089	80 0,045	83 0,025	87 0,0099	90 0,0045
	9	79 0,186	84 0,096	89 0,043	93 0,021	97 0,0089	100 0,0043
	10	87 0,186	93 0,088	97 0,049	102 0,021	106 0,0099	110 0,0041
	11	95 0,186	101 0,093	106 0,047	111 0,022	116 0,0088	119 0,0048
	12	103 0,186	109 0,098	115 0,046	120 0,022	125 0,0096	129 0,0045
	13	111 0,186	118 0,091	124 0,044	129 0,022	135 0,0086	139 0,0042
14	2	23 0,200	25 0,100	27 0,033	28 0,017		
	3	32 0,197	35 0,091	37 0,047	39 0,021	41 0,0059	42 0,0029
	4	41 0,192	45 0,079	47 0,046	49 0,025	52 0,0078	53 0,0046
	5	50 0,186	54 0,087	57 0,044	60 0,019	63 0,0072	64 0,0050
	6	59 0,179	63 0,091	67 0,041	70 0,020	73 0,0087	75 0,0046
	7	67 0,197	72 0,094	76 0,046	79 0,025	83 0,0097	86 0,0042
	8	76 0,188	81 0,095	86 0,042	89 0,024	94 0,0081	96 0,0050
	9	85 0,179	90 0,096	95 0,046	99 0,023	104 0,0086	107 0,0043
	10	93 0,192	99 0,096	104 0,048	109 0,022	114 0,0089	117 0,0048
	11	102 0,183	108 0,095	114 0,044	118 0,025	124 0,0090	128 0,0042
	12	110 0,193	117 0,095	123 0,046	128 0,023	134 0,0091	138 0,0045
	13	119 0,185	126 0,095	132 0,048	138 0,022	144 0,0091	148 0,0047
	14	127 0,194	135 0,094	141 0,050	147 0,024	154 0,0091	158 0,0049
15	2	25 0,177	27 0,088	29 0,029	30 0,015		
	3	35 0,164	35 0,076	40 0,039	42 0,017	43 0,0098	44 0,0049
	4	44 0,185	48 0,080	50 0,049	53 0,020	55 0,0093	57 0,0036

TABLA 6 Valores críticos y *P* valores de U_s para el test de Wilcoxon-Mann-Whitney (*continuación*)

<i>n</i>	<i>n'</i>	0,20	0,10	0,05	0,025	0,01	0,005
16	5	53 0,197	57 0,098	61 0,042	64 0,019	67 0,0077	69 0,0037
	6	63 0,178	67 0,095	71 0,045	74 0,023	78 0,0084	80 0,0046
	7	72 0,185	77 0,091	81 0,047	85 0,021	89 0,0085	92 0,0038
	8	81 0,190	87 0,087	91 0,047	95 0,024	100 0,0085	103 0,0042
	9	90 0,194	96 0,096	101 0,048	106 0,021	111 0,0083	114 0,0044
	10	99 0,196	106 0,091	111 0,048	116 0,023	121 0,0096	125 0,0044
	11	108 0,198	115 0,097	121 0,047	126 0,024	132 0,0092	136 0,0045
	12	117 0,200	125 0,093	131 0,047	136 0,025	143 0,0087	147 0,0044
	13	127 0,185	134 0,098	141 0,046	147 0,022	153 0,0096	158 0,0044
	14	136 0,186	144 0,093	151 0,046	157 0,023	164 0,0091	169 0,0043
	15	145 0,187	153 0,098	161 0,045	167 0,024	174 0,0099	179 0,0049
	2	27 0,157	29 0,078	31 0,026	32 0,013		
	3	37 0,171	40 0,085	42 0,048	44 0,023	46 0,0083	47 0,0041
	4	47 0,178	50 0,100	53 0,050	56 0,022	59 0,0074	60 0,0050
	5	57 0,179	61 0,091	65 0,040	67 0,025	71 0,0082	73 0,0041
17	6	67 0,178	71 0,098	75 0,049	79 0,021	83 0,0080	85 0,0045
	7	76 0,198	82 0,089	86 0,047	90 0,023	94 0,0096	97 0,0046
	8	86 0,192	92 0,093	97 0,045	101 0,023	106 0,0087	109 0,0045
	9	96 0,187	102 0,095	107 0,049	112 0,023	117 0,0096	121 0,0043
	10	106 0,182	112 0,097	118 0,047	123 0,023	129 0,0087	133 0,0041
	11	115 0,195	122 0,099	129 0,044	134 0,023	140 0,0093	144 0,0047
	12	125 0,189	132 0,047	139 0,047	145 0,023	151 0,0097	156 0,0044
	13	134 0,199	143 0,092	149 0,050	156 0,022	163 0,0087	167 0,0048
	14	144 0,193	153 0,093	160 0,047	166 0,025	174 0,0091	179 0,0045
	15	154 0,188	163 0,093	170 0,049	177 0,024	185 0,0093	190 0,0048
	16	163 0,196	173 0,094	181 0,047	188 0,023	196 0,0096	202 0,0045
	2	28 0,187	31 0,070	32 0,047	33 0,023		
	3	39 0,179	42 0,093	45 0,040	47 0,019	49 0,0070	50 0,0035
	4	50 0,172	53 0,099	57 0,040	59 0,024	62 0,0090	64 0,0040
	5	60 0,189	65 0,085	68 0,048	71 0,025	75 0,0086	77 0,0046
18	6	71 0,177	76 0,087	80 0,044	83 0,024	87 0,0099	90 0,0045
	7	81 0,187	86 0,100	91 0,047	95 0,024	100 0,0085	103 0,0042
	8	91 0,194	97 0,098	102 0,050	107 0,023	112 0,0090	115 0,0048
	9	101 0,200	108 0,095	114 0,045	118 0,025	124 0,0092	128 0,0043
	10	112 0,187	119 0,093	125 0,046	130 0,024	136 0,0093	140 0,0047
	11	122 0,191	130 0,091	136 0,047	142 0,022	136 0,0093	152 0,0049
	12	132 0,195	140 0,097	147 0,048	153 0,024	160 0,0093	165 0,0043
	13	142 0,198	151 0,095	158 0,048	164 0,025	172 0,0091	177 0,0045
	14	153 0,186	161 0,100	169 0,048	176 0,023	184 0,0090	189 0,0046
	15	163 0,189	172 0,097	180 0,049	187 0,024	195 0,0100	201 0,0047
	16	173 0,191	183 0,094	191 0,049	199 0,023	207 0,0097	213 0,0048
	17	183 0,193	193 0,099	202 0,049	210 0,024	219 0,0095	225 0,0048
	2	30 0,168	32 0,095	34 0,042	35 0,021		
	3	41 0,185	45 0,080	47 0,047	49 0,024	52 0,0060	53 0,0030
	4	52 0,195	56 0,098	60 0,042	63 0,019	66 0,0074	67 0,0049
	5	63 0,200	68 0,094	72 0,046	75 0,024	79 0,0089	81 0,0049
	6	74 0,199	80 0,090	84 0,047	88 0,022	92 0,0094	95 0,0044
	7	85 0,198	91 0,097	96 0,047	100 0,025	105 0,0094	108 0,0049
	8	96 0,196	103 0,091	108 0,047	113 0,022	118 0,0092	122 0,0042

TABLA 6 Valores críticos y *P* valores de U_s para el test de Wilcoxon-Mann-Whitney (*continuación*)

<i>n</i>	<i>n'</i>	0,20	0,10	0,05	0,025	0,01	0,005
19	9	107 0,194	114 0,095	120 0,046	125 0,023	131 0,0089	135 0,0043
	10	118 0,191	125 0,099	132 0,045	137 0,024	143 0,0100	148 0,0044
	11	129 0,188	137 0,092	143 0,049	149 0,024	156 0,0094	161 0,0043
	12	139 0,200	148 0,095	155 0,048	161 0,025	169 0,0089	173 0,0050
	13	150 0,196	159 0,097	167 0,046	173 0,025	181 0,0095	186 0,0049
	14	161 0,193	170 0,099	178 0,049	185 0,025	194 0,0089	199 0,0047
	15	172 0,190	182 0,093	190 0,048	197 0,025	206 0,0094	212 0,0046
	16	182 0,199	193 0,095	202 0,046	209 0,025	218 0,0099	224 0,0050
	17	193 0,195	204 0,096	213 0,049	221 0,025	231 0,0093	237 0,0048
	18	204 0,192	215 0,097	225 0,047	233 0,024	243 0,0096	250 0,0046
	2	31 0,191	34 0,086	36 0,038	37 0,019	38 0,0095	
	3	43 0,191	47 0,087	50 0,040	52 0,021	54 0,0091	56 0,0026
	4	55 0,188	59 0,097	63 0,044	66 0,021	69 0,0086	71 0,0041
	5	67 0,183	72 0,088	76 0,044	79 0,024	83 0,0093	86 0,0039
	6	78 0,198	84 0,092	89 0,043	93 0,021	97 0,0090	100 0,0044
	7	90 0,188	96 0,094	101 0,048	106 0,022	111 0,0085	114 0,0045
	8	101 0,198	108 0,095	114 0,045	119 0,022	124 0,0094	128 0,0044
	9	113 0,188	120 0,095	126 0,048	131 0,025	138 0,0086	142 0,0043
	10	124 0,195	132 0,094	138 0,050	144 0,024	151 0,0091	155 0,0048
20	11	136 0,185	144 0,094	151 0,047	157 0,023	164 0,0094	169 0,0045
	12	147 0,191	156 0,093	163 0,048	170 0,023	177 0,0097	182 0,0049
	13	158 0,195	167 0,100	175 0,049	182 0,025	190 0,0098	196 0,0045
	14	169 0,199	179 0,098	188 0,046	195 0,024	203 0,0099	209 0,0048
	15	181 0,190	191 0,096	200 0,047	208 0,023	216 0,0100	223 0,0045
	16	192 0,194	203 0,095	212 0,048	220 0,024	230 0,0090	236 0,0047
	17	203 0,196	214 0,100	224 0,049	233 0,023	242 0,0100	249 0,0048
	18	214 0,199	226 0,098	236 0,049	245 0,024	255 0,0100	262 0,0050
	19	226 0,191	238 0,096	248 0,050	258 0,023	268 0,0099	276 0,0046
	2	33 0,173	36 0,078	38 0,035	39 0,017	40 0,0087	
	3	45 0,197	49 0,094	52 0,046	55 0,018	57 0,0079	58 0,0045
	4	58 0,183	62 0,097	66 0,045	69 0,023	72 0,0100	75 0,0034
	5	70 0,192	75 0,097	80 0,042	83 0,024	87 0,0096	90 0,0043
	6	82 0,196	88 0,095	93 0,046	97 0,023	102 0,0087	105 0,0043
	7	94 0,198	101 0,092	106 0,048	111 0,022	116 0,0093	120 0,0041
	8	106 0,199	113 0,099	119 0,049	124 0,025	130 0,0096	134 0,0047
	9	118 0,199	126 0,095	132 0,049	138 0,023	144 0,0097	149 0,0043
	10	130 0,198	138 0,100	145 0,049	151 0,024	158 0,0096	163 0,0045
	11	142 0,197	151 0,095	158 0,049	165 0,023	172 0,0095	177 0,0047
	12	154 0,195	163 0,099	171 0,048	178 0,024	186 0,0092	191 0,0048
	13	166 0,194	176 0,094	184 0,048	191 0,024	200 0,0090	205 0,0049
	14	178 0,192	188 0,097	197 0,047	204 0,025	213 0,0098	219 0,0049
	15	190 0,191	200 0,099	210 0,046	218 0,023	227 0,0095	233 0,0049
	16	201 0,200	213 0,095	222 0,049	231 0,024	241 0,0091	247 0,0049
	17	213 0,198	225 0,097	235 0,049	244 0,024	254 0,0097	261 0,0048
	18	225 0,196	237 0,099	248 0,048	257 0,024	268 0,0094	275 0,0048
	19	237 0,194	250 0,095	261 0,047	270 0,024	281 0,0099	289 0,0047
	20	249 0,192	262 0,097	273 0,049	283 0,025	295 0,0095	303 0,0047

TABLA 7 Valores críticos y P valores de B_s para el test de signos

Nota: debido a que la distribución nula del test de signos es discreta, esta tabla proporciona valores seleccionados del estadístico de contraste B_s **en negrilla** y los correspondientes P valores para una alternativa no direccional *en cursiva*. Los P valores direccionales se obtienen dividiendo por dos los números en cursiva.

n_d	0,20	0,10	0,05	0,02	0,01	0,002	0,001
1							
2							
3							
4							
5	5 0,063	5 0,063					
6	6 0,031	6 0,031	6 0,031				
7	6 0,125	7 0,016	7 0,016	7 0,016			
8	7 0,070	7 0,070	8 0,008	8 0,008	8 0,008		
9	7 0,180	8 0,039	8 0,039	9 0,004	9 0,004		
10	8 0,109	9 0,021	9 0,021	10 0,002	10 0,002	10 0,0020	
11	9 0,065	9 0,065	10 0,012	10 0,012	11 0,001	11 0,0010	11 0,0010
12	9 0,146	10 0,039	10 0,039	11 0,006	11 0,006	12 0,0005	12 0,0005
13	10 0,092	10 0,093	11 0,023	12 0,003	12 0,003	13 0,0002	13 0,0002
14	10 0,180	11 0,057	12 0,013	12 0,013	13 0,0018	13 0,0018	14 0,0001
15	11 0,118	12 0,035	12 0,035	13 0,007	13 0,007	14 0,0010	14 0,0010
16	12 0,077	12 0,077	13 0,021	14 0,004	14 0,004	15 0,0005	15 0,0005
17	12 0,143	13 0,049	13 0,049	14 0,013	15 0,002	16 0,0003	16 0,0003
18	13 0,096	13 0,096	14 0,031	15 0,008	15 0,008	16 0,0013	17 0,0001
19	13 0,167	14 0,064	15 0,019	15 0,019	16 0,004	17 0,0007	17 0,0007
20	14 0,115	15 0,041	15 0,041	16 0,012	17 0,003	18 0,0004	18 0,0004
21	14 0,189	15 0,078	16 0,027	17 0,007	17 0,007	18 0,0015	19 0,0002
22	15 0,134	16 0,052	17 0,017	17 0,017	18 0,004	19 0,0009	19 0,0009
23	16 0,093	16 0,093	17 0,037	18 0,011	19 0,003	20 0,0005	20 0,0005
24	16 0,152	17 0,064	18 0,023	19 0,007	19 0,007	20 0,0015	21 0,0003
25	17 0,108	18 0,043	18 0,043	19 0,015	20 0,004	21 0,0009	21 0,0009
26	17 0,168	18 0,076	19 0,029	20 0,009	20 0,009	22 0,0005	22 0,0005
27	18 0,122	19 0,052	20 0,019	20 0,019	21 0,006	22 0,0015	23 0,0003
28	18 0,185	19 0,087	20 0,036	21 0,013	22 0,004	23 0,0009	23 0,0009
29	19 0,136	20 0,061	21 0,024	22 0,008	22 0,008	24 0,0005	24 0,0005
30	20 0,099	20 0,099	21 0,043	22 0,016	23 0,005	24 0,0014	25 0,0003
31	20 0,152	21 0,071	22 0,029	23 0,011	24 0,003	25 0,0009	25 0,0009

TABLA 8 Valores críticos de P valores de W_s para el test de rangos con signo de Wilcoxon

Nota: debido a que la distribución nula del test de rangos con signo de Wilcoxon es discreta, esta tabla proporciona valores seleccionados del estadístico de contraste W_s **en negrilla** y los correspondientes P valores para una alternativa no direccional *en cursiva*. Los P valores direccionales se obtienen dividiendo por dos los números en cursiva.

n	0,20	0,10	0,05	0,02	0,01	0,002	0,001
1							
2							
3							
4	10 0,125						
5	13 0,188	15 0,063					
6	18 0,156	19 0,093	21 0,031				
7	23 0,156	25 0,078	26 0,047	28 0,016			
8	28 0,195	31 0,078	33 0,039	35 0,016	36 0,0078		
9	35 0,164	37 0,098	40 0,039	42 0,020	44 0,0078		
10	41 0,193	45 0,084	47 0,049	50 0,020	52 0,0098	55 0,0020	
11	49 0,175	53 0,083	56 0,042	59 0,019	61 0,0098	65 0,0020	66 0,0010
12	57 0,176	61 0,092	65 0,042	69 0,016	71 0,0093	76 0,0015	77 0,0010
13	65 0,191	70 0,094	74 0,048	79 0,017	82 0,0081	87 0,0017	89 0,0007
14	74 0,194	80 0,091	84 0,049	90 0,017	93 0,0085	99 0,0017	101 0,0009
15	84 0,188	90 0,095	95 0,048	101 0,018	105 0,0084	112 0,0015	114 0,0009
16	94 0,193	101 0,093	107 0,044	113 0,018	117 0,0092	125 0,0017	128 0,0008
17	105 0,190	112 0,098	119 0,045	126 0,017	130 0,0093	139 0,0017	142 0,0008
18	116 0,196	124 0,099	131 0,048	139 0,018	144 0,0090	153 0,0019	157 0,0008
19	128 0,196	137 0,096	144 0,049	153 0,018	158 0,0094	169 0,0017	172 0,0010
20	141 0,189	150 0,097	158 0,048	167 0,019	173 0,0094	184 0,0020	189 0,0009
21	154 0,191	164 0,096	173 0,046	182 0,019	189 0,0090	201 0,0019	206 0,0009
22	167 0,198	178 0,094	188 0,046	198 0,019	205 0,0093	218 0,0019	223 0,0009
23	182 0,190	193 0,098	203 0,048	214 0,020	222 0,0091	236 0,0019	241 0,0010
24	196 0,197	209 0,095	219 0,049	231 0,019	239 0,0096	255 0,0018	260 0,0010
25	212 0,191	225 0,096	236 0,048	249 0,019	257 0,0096	274 0,0018	280 0,0009
26	227 0,199	241 0,099	253 0,049	267 0,019	276 0,0094	293 0,0020	300 0,0009
27	244 0,194	259 0,095	271 0,049	286 0,019	295 0,0096	314 0,0019	321 0,0009
28	261 0,194	276 0,099	290 0,048	305 0,019	315 0,0095	335 0,0019	342 0,0010
29	278 0,198	295 0,096	309 0,048	325 0,019	335 0,0099	256 0,0020	364 0,0010
30	296 0,198	314 0,096	328 0,050	345 0,020	356 0,0099	379 0,0019	387 0,0010
31	315 0,195	333 0,098	349 0,048	366 0,020	378 0,0097	402 0,0019	410 0,0010

TABLA 9 Valores críticos de la distribución chi-cuadrado

Nota: las cabeceras de las columnas son los P valores no direccionales (omnidireccionales).

Si H_A es direccional (lo que solo es posible cuando $gl = 1$), los P valores direccionales se obtienen dividiendo las cabeceras de las columnas por dos.

gl	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	1,64	2,71	3,84	5,41	6,63	10,83	15,14
2	3,22	4,61	5,99	7,82	9,21	13,82	18,42
3	4,64	6,25	7,81	9,84	11,34	16,27	21,11
4	5,99	7,78	9,49	11,67	13,28	18,47	23,51
5	7,29	9,24	11,07	13,39	15,09	20,51	25,74
6	8,56	10,64	12,59	15,03	16,81	22,46	27,86
7	9,80	12,02	14,07	16,62	18,48	24,32	29,88
8	11,03	13,36	15,51	18,17	20,09	26,12	31,83
9	12,24	14,68	16,92	19,68	21,67	27,88	33,72
10	13,44	15,99	18,31	21,16	23,21	29,59	35,56
11	14,63	17,28	19,68	22,62	24,72	31,26	37,37
12	15,81	18,55	21,03	24,05	26,22	32,91	39,13
13	16,98	19,81	22,36	25,47	27,69	34,53	40,87
14	18,15	21,06	23,68	26,87	29,14	36,12	42,58
15	19,31	22,31	25,00	28,26	30,58	37,70	44,26
16	20,47	23,54	26,30	29,63	32,00	39,25	45,92
17	21,61	24,77	27,59	31,00	33,41	40,79	47,57
18	22,76	25,99	28,87	32,35	34,81	42,31	49,19
19	23,90	27,20	30,14	33,69	36,19	43,82	50,80
20	25,04	28,41	31,41	35,02	37,57	45,31	52,39
21	26,17	29,62	32,67	36,34	38,93	46,80	53,96
22	27,30	30,81	33,92	37,66	40,29	48,27	55,52
23	28,43	32,01	35,17	38,97	41,64	49,73	57,08
24	29,55	33,20	36,42	40,27	42,98	51,18	58,61
25	30,68	34,38	37,65	41,57	44,31	52,62	60,14
26	31,79	35,56	38,89	42,86	45,64	54,05	61,66
27	32,91	36,74	40,11	44,14	46,96	55,48	63,16
28	34,03	37,92	41,34	45,42	48,28	56,89	64,66
29	35,14	39,09	42,56	46,69	49,59	58,30	66,15
30	36,25	40,26	43,77	47,96	50,89	59,70	67,63

TABLA 10 Valores críticos de la distribución F
gl del numerador = 1

gl del denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	9,47	39,86	161	101 ¹	405 ¹	406 ³	405 ⁵
2	3,56	8,53	18,51	48,51	98,50	998	100 ²
3	2,68	5,54	10,13	20,62	34,12	167	784
4	2,35	4,54	7,71	14,04	21,20	74,14	242
5	2,18	4,06	6,61	11,32	16,26	47,18	125
6	2,07	3,78	5,99	9,88	13,75	35,51	82,49
7	2,00	3,59	5,59	8,99	12,25	29,25	62,17
8	1,95	3,46	5,32	8,39	11,26	25,41	50,69
9	1,91	3,36	5,12	7,96	10,56	22,86	43,48
10	1,88	3,29	4,96	7,64	10,04	21,04	38,58
11	1,86	3,23	4,84	7,39	9,65	19,69	35,06
12	1,84	3,18	4,75	7,19	9,33	18,64	32,43
13	1,82	3,14	4,67	7,02	9,07	17,82	30,39
14	1,81	3,10	4,60	6,89	8,86	17,14	28,77
15	1,80	3,07	4,54	6,77	8,68	16,59	27,45
16	1,79	3,05	4,49	6,67	8,53	16,12	26,36
17	1,78	3,03	4,45	6,59	8,40	15,72	25,44
18	1,77	3,01	4,41	6,51	8,29	15,38	24,66
19	1,76	2,99	4,38	6,45	8,18	15,08	23,99
20	1,76	2,97	4,35	6,39	8,10	14,82	23,40
21	1,75	2,96	4,32	6,34	8,02	14,59	22,89
22	1,75	2,95	4,30	6,29	7,95	14,38	22,43
23	1,74	2,94	4,28	6,25	7,88	14,20	22,03
24	1,74	2,93	4,26	6,21	7,82	14,03	21,66
25	1,73	2,92	4,24	6,18	7,77	13,88	21,34
26	1,73	2,91	4,23	6,14	7,72	13,74	21,04
27	1,73	2,90	4,21	6,11	7,68	13,61	20,77
28	1,72	2,89	4,20	6,09	7,64	13,50	20,53
29	1,72	2,89	4,18	6,06	7,60	13,39	20,30
30	1,72	2,88	4,17	6,04	7,56	13,29	20,09
40	1,70	2,84	4,08	5,87	7,31	12,61	18,67
60	1,68	2,79	4,00	5,71	7,08	11,97	17,38
100	1,66	2,76	3,94	5,59	6,90	11,50	16,43
140	1,66	2,74	3,91	5,54	6,82	11,30	16,05
∞	1,64	2,71	3,84	5,41	6,63	10,83	15,14

Notación: 406³ significa 406×10^3 .

(Continúa)

TABLA 10 Valores críticos de la distribución *F* (continuación)
gl del numerador = 2

gl del denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	12,00	49,50	200	125 ¹	500 ¹	500 ³	500 ⁵
2	4,00	9,00	19,00	49,00	99,00	999	100 ²
3	2,89	5,46	9,55	18,86	30,82	149	695
4	2,47	4,32	6,94	12,14	18,00	61,25	198
5	2,26	3,78	5,79	9,45	13,27	37,12	97,03
6	2,13	3,46	5,14	8,05	10,92	27,00	61,63
7	2,04	3,26	4,74	7,20	9,55	21,69	45,13
8	1,98	3,11	4,46	6,64	8,65	18,49	36,00
9	1,93	3,01	4,26	6,23	8,02	16,39	30,34
10	1,90	2,92	4,10	5,93	7,56	14,91	26,55
11	1,87	2,86	3,98	5,70	7,21	13,81	23,85
12	1,85	2,81	3,89	5,52	6,93	12,97	21,85
13	1,83	2,76	3,81	5,37	6,70	12,31	20,31
14	1,81	2,73	3,74	5,24	6,51	11,78	19,09
15	1,80	2,70	3,68	5,14	6,36	11,34	18,11
16	1,78	2,67	3,63	5,05	6,23	10,97	17,30
17	1,77	2,64	3,59	4,97	6,11	10,66	16,62
18	1,76	2,62	3,55	4,90	6,01	10,39	16,04
19	1,75	2,61	3,52	4,84	5,93	10,16	15,55
20	1,75	2,59	3,49	4,79	5,85	9,95	15,12
21	1,74	2,57	3,47	4,74	5,78	9,77	14,74
22	1,73	2,56	3,44	4,70	5,72	9,61	14,41
23	1,73	2,55	3,42	4,66	5,66	9,47	14,12
24	1,72	2,54	3,40	4,63	5,61	9,34	13,85
25	1,72	2,53	3,39	4,59	5,57	9,22	13,62
26	1,71	2,52	3,37	4,56	5,53	9,12	13,40
27	1,71	2,51	3,35	4,54	5,49	9,02	13,21
28	1,71	2,50	3,34	4,51	5,45	8,93	13,03
29	1,70	2,50	3,33	4,49	5,42	8,85	12,87
30	1,70	2,49	3,32	4,47	5,39	8,77	12,72
40	1,68	2,44	3,23	4,32	5,18	8,25	11,70
60	1,65	2,39	3,15	4,18	4,98	7,77	10,78
100	1,64	2,36	3,09	4,07	4,82	7,41	10,11
140	1,63	2,34	3,06	4,02	4,76	7,26	9,84
∞	1,61	2,30	3,00	3,91	4,61	6,91	9,21

TABLA 10 Valores críticos de la distribución *F* (continuación)

gl del numerador = 3

gl del denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	13,06	53,59	216	135 ¹	540 ¹	540 ³	540 ⁵
2	4,16	9,16	19,16	49,17	99,17	999	100 ²
3	2,94	5,39	9,28	18,11	29,46	141	659
4	2,48	4,19	6,59	11,34	16,69	56,18	181
5	2,25	3,62	5,41	8,67	12,06	33,20	86,29
6	2,11	3,29	4,76	7,29	9,78	23,70	53,68
7	2,02	3,07	4,35	6,45	8,45	18,77	38,68
8	1,95	2,92	4,07	5,90	7,59	15,83	30,46
9	1,90	2,81	3,86	5,51	6,99	13,90	25,40
10	1,86	2,73	3,71	5,22	6,55	12,55	22,04
11	1,83	2,66	3,59	4,99	6,22	11,56	19,66
12	1,80	2,61	3,49	4,81	5,95	10,80	17,90
13	1,78	2,56	3,41	4,67	5,74	10,21	16,55
14	1,76	2,52	3,34	4,55	5,56	9,73	15,49
15	1,75	2,49	3,29	4,45	5,42	9,34	14,64
16	1,74	2,46	3,24	4,36	5,29	9,01	13,93
17	1,72	2,44	3,20	4,29	5,18	9,73	13,34
18	1,71	2,42	3,16	4,22	5,09	8,49	12,85
19	1,70	2,40	3,13	4,16	5,01	8,28	12,42
20	1,70	2,38	3,10	4,11	4,94	8,10	12,05
21	1,69	2,36	3,07	4,07	4,87	7,94	11,73
22	1,68	2,35	3,05	4,03	4,82	7,80	11,44
23	1,68	2,34	3,03	3,99	4,76	7,67	11,19
24	1,67	2,33	3,01	3,96	4,72	7,55	10,96
25	1,66	2,32	2,99	3,93	4,68	7,45	10,76
26	1,66	2,31	2,98	3,90	4,64	7,36	10,58
27	1,65	2,30	2,96	3,87	4,60	7,27	10,41
28	1,65	2,29	2,95	3,85	4,57	7,19	10,26
29	1,65	2,28	2,93	3,83	4,54	7,12	10,12
30	1,64	2,28	2,92	3,81	4,51	7,05	9,99
40	1,62	2,23	2,84	3,67	4,31	6,59	9,13
60	1,60	2,18	2,76	3,53	4,13	6,17	8,35
100	1,58	2,14	2,70	3,43	3,98	5,86	7,79
140	1,57	2,12	2,67	3,38	3,92	5,73	7,57
∞	1,55	2,08	2,60	3,28	3,78	5,42	7,04

TABLA 10 Valores críticos de la distribución F (continuación)
gl del numerador = 4

gl del denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	13,64	55,83	225	141 ¹	562 ¹	562 ³	562 ⁵
2	4,24	9,24	19,25	49,25	99,25	999	100 ²
3	2,96	5,34	9,12	17,69	28,71	137	640
4	2,48	4,11	6,39	10,90	15,98	53,44	172
5	2,24	3,52	5,19	8,23	11,39	31,09	80,53
6	2,09	3,18	4,53	6,86	9,15	21,92	49,42
7	1,99	2,96	4,12	6,03	7,85	17,20	35,22
8	1,92	2,81	3,84	5,49	7,01	14,39	27,49
9	1,87	2,69	3,63	5,10	6,42	12,56	22,77
10	1,83	2,61	3,48	4,82	5,99	11,28	19,63
11	1,80	2,54	3,36	4,59	5,67	10,35	17,42
12	1,77	2,48	3,26	4,42	5,41	9,63	15,79
13	1,75	2,43	3,18	4,28	5,21	9,07	14,55
14	1,73	2,39	3,11	4,16	5,04	8,62	13,57
15	1,71	2,36	3,06	4,06	4,89	8,25	12,78
16	1,70	2,33	3,01	3,97	4,77	7,94	12,14
17	1,68	2,31	2,96	3,90	4,67	7,68	11,60
18	1,67	2,29	2,93	3,84	4,58	7,46	11,14
19	1,66	2,27	2,90	3,78	4,50	7,27	10,75
20	1,65	2,25	2,87	3,73	4,43	7,10	10,41
21	1,65	2,23	2,84	3,69	4,37	6,95	10,12
22	1,64	2,22	2,82	3,65	4,31	6,81	9,86
23	1,63	2,21	2,80	3,61	4,26	6,70	9,63
24	1,63	2,19	2,78	3,58	4,22	6,59	9,42
25	1,62	2,18	2,76	3,55	4,18	6,49	9,24
26	1,62	2,17	2,74	3,52	4,14	6,41	9,07
27	1,61	2,17	2,73	3,50	4,11	6,33	8,92
28	1,61	2,16	2,71	3,47	4,07	6,25	8,79
29	1,60	2,15	2,70	3,45	4,04	6,19	8,66
30	1,60	2,14	2,69	3,43	4,02	6,12	8,54
40	1,57	2,09	2,61	3,30	3,83	5,70	7,76
60	1,55	2,04	2,53	3,16	3,65	5,31	7,06
100	1,53	2,00	2,46	3,06	3,51	5,02	6,55
140	1,52	1,99	2,44	3,02	3,46	4,90	6,35
∞	1,50	1,94	2,37	2,92	3,32	4,62	5,88

TABLA 10 Valores críticos de la distribución *F* (continuación)

gl del numerador = 5

gl del denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	14,01	52,74	230	144 ¹	576 ¹	576 ³	576 ⁵
2	4,28	9,29	19,30	49,30	99,30	999	100 ²
3	2,97	5,31	9,01	17,43	28,24	135	628
4	2,48	4,05	6,26	10,62	15,52	51,71	166
5	2,23	3,45	5,05	7,95	10,97	29,75	76,91
6	2,08	3,11	4,39	6,58	8,75	20,80	46,75
7	1,97	2,88	3,97	5,76	7,46	16,21	33,06
8	1,90	2,73	3,69	5,22	6,63	13,48	25,63
9	1,85	2,61	3,48	4,84	6,06	11,71	21,11
10	1,80	2,52	3,33	4,55	5,64	10,48	18,12
11	1,77	2,45	3,20	4,34	5,32	9,58	16,02
12	1,74	2,39	3,11	4,16	5,06	8,89	14,47
13	1,72	2,35	3,03	4,02	4,86	8,35	13,29
14	1,70	2,31	2,96	3,90	4,69	7,92	12,37
15	1,68	2,27	2,90	3,81	4,56	7,57	11,62
16	1,67	2,24	2,85	3,72	4,44	7,27	11,01
17	1,65	2,22	2,81	3,65	4,34	7,02	10,50
18	1,64	2,20	2,77	3,59	4,25	6,81	10,07
19	1,63	2,18	2,74	3,53	4,17	6,62	9,71
20	1,62	2,16	2,71	3,48	4,10	6,46	9,39
21	1,61	2,14	2,68	3,44	4,04	6,32	9,11
22	1,61	2,13	2,66	3,40	3,99	6,19	8,87
23	1,60	2,11	2,64	3,36	3,94	6,08	8,65
24	1,59	2,10	2,62	3,33	3,90	5,98	8,46
25	1,59	2,09	2,60	3,30	3,85	5,89	8,28
26	1,58	2,08	2,59	3,28	3,82	5,80	8,13
27	1,58	2,07	2,57	3,25	3,78	5,73	7,99
28	1,57	2,06	2,56	3,23	3,75	5,66	7,86
29	1,57	2,06	2,55	3,21	3,73	5,59	7,74
30	1,57	2,05	2,53	3,19	3,70	5,53	7,63
40	1,54	2,00	2,45	3,05	3,51	5,13	6,90
60	1,51	1,95	2,37	2,92	3,34	4,76	6,25
100	1,49	1,91	2,31	2,82	3,21	4,48	5,78
140	1,48	1,89	2,28	2,78	3,15	4,37	5,59
∞	1,46	1,85	2,21	2,68	3,02	4,10	5,15

TABLA 10 Valores críticos de la distribución F (continuación)
gl del numerador = 6

gl del. denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	14,26	58,20	234	146 ¹	586 ¹	586 ³	586 ⁵
2	4,32	9,33	19,33	49,33	99,33	999	100 ²
3	2,97	5,28	8,94	17,25	27,91	133	620
4	2,47	4,01	6,16	10,42	15,21	50,53	162
5	2,22	3,40	4,95	7,76	10,67	28,83	74,43
6	2,06	3,05	4,28	6,39	8,47	20,03	44,91
7	1,96	2,83	3,87	5,58	7,19	15,52	31,57
8	1,88	2,67	3,58	5,04	6,37	12,86	24,36
9	1,83	2,55	3,37	4,65	5,80	11,13	19,97
10	1,78	2,46	3,22	4,37	5,39	9,93	17,08
11	1,75	2,39	3,09	4,15	5,07	9,05	15,05
12	1,72	2,33	3,00	3,98	4,82	8,38	13,56
13	1,69	2,28	2,92	3,84	4,62	7,86	12,42
14	1,67	2,24	2,85	3,72	4,46	7,44	11,53
15	1,66	2,21	2,79	3,63	4,32	7,09	10,82
16	1,64	2,18	2,74	3,54	4,20	6,80	10,23
17	1,63	2,15	2,70	3,47	4,10	6,56	9,75
18	1,62	2,13	2,66	3,41	4,01	6,35	9,33
19	1,61	2,11	2,63	3,35	3,94	6,18	8,98
20	1,60	2,09	2,60	3,30	3,87	6,02	8,68
21	1,59	2,08	2,57	3,26	3,81	5,88	8,41
22	1,58	2,06	2,55	3,22	3,76	5,76	8,18
23	1,57	2,05	2,53	3,19	3,71	5,65	7,97
24	1,57	2,04	2,51	3,15	3,67	5,55	7,79
25	1,56	2,02	2,49	3,13	3,63	5,46	7,62
26	1,56	2,01	2,47	3,10	3,59	5,38	7,48
27	1,55	2,00	2,46	3,07	3,56	5,31	7,34
28	1,55	2,00	2,45	3,05	3,53	5,24	7,22
29	1,54	1,99	2,43	3,03	3,50	5,18	7,10
30	1,54	1,98	2,42	3,01	3,47	5,12	7,00
40	1,51	1,93	2,34	2,88	3,29	4,73	6,30
60	1,48	1,87	2,25	2,75	3,12	4,37	5,68
100	1,46	1,83	2,19	2,65	2,99	4,11	5,24
140	1,45	1,82	2,16	2,61	2,93	4,00	5,06
∞	1,43	1,77	2,10	2,51	2,80	3,74	4,64

TABLA 10 Valores críticos de la distribución *F* (continuación)

gl del numerador = 7

gl denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	14,44	58,91	237	148 ¹	593 ¹	593 ³	593 ⁵
2	4,34	9,35	19,35	49,36	99,36	999	100 ²
3	2,97	5,27	8,89	17,11	27,67	132	614
4	2,47	3,98	6,09	10,27	14,98	49,66	159
5	2,21	3,37	4,88	7,61	10,46	28,16	72,61
6	2,05	3,01	4,21	6,25	8,26	19,46	43,57
7	1,94	2,78	3,79	5,44	6,99	15,02	30,48
8	1,87	2,62	3,50	4,90	6,18	12,40	23,42
9	1,81	2,51	3,29	4,52	5,61	10,70	19,14
10	1,77	2,41	3,14	4,23	5,20	9,52	16,32
11	1,73	2,34	3,01	4,02	4,89	8,66	14,34
12	1,70	2,28	2,91	3,85	4,64	8,00	12,89
13	1,68	2,23	2,83	3,71	4,44	7,49	11,79
14	1,65	2,19	2,76	3,59	4,28	7,08	10,92
15	1,64	2,16	2,71	3,49	4,14	6,74	10,23
16	1,62	2,13	2,66	3,41	4,03	6,46	9,66
17	1,61	2,10	2,61	3,34	3,93	6,22	9,19
18	1,60	2,08	2,58	3,27	3,84	6,02	8,79
19	1,58	2,06	2,54	3,22	3,77	5,85	8,45
20	1,58	2,04	2,51	3,17	3,70	5,69	8,16
21	1,57	2,02	2,49	3,13	3,64	5,56	7,90
22	1,56	2,01	2,46	3,09	3,59	5,44	7,68
23	1,55	1,99	2,44	3,05	3,54	5,33	7,48
24	1,55	1,98	2,42	3,02	3,50	5,23	7,30
25	1,54	1,97	2,40	2,99	3,46	5,15	7,14
26	1,53	1,96	2,39	2,97	3,42	5,07	6,99
27	1,53	1,95	2,37	2,94	3,39	5,00	6,86
28	1,52	1,94	2,36	2,92	3,36	4,93	6,75
29	1,52	1,93	2,35	2,90	3,33	4,87	6,64
30	1,52	1,93	2,33	2,88	3,30	4,82	6,54
40	1,49	1,87	2,25	2,74	3,12	4,44	5,86
60	1,46	1,82	2,17	2,62	2,95	4,09	5,27
100	1,43	1,78	2,10	2,52	2,82	3,83	4,84
140	1,42	1,76	2,08	2,48	2,77	3,72	4,67
∞	1,40	1,72	2,01	2,37	2,64	3,47	4,27

TABLA 10 Valores críticos de la distribución F (continuación)
gl del numerador = 8

gl del denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	14,58	59,44	239	149 ¹	598 ¹	598 ³	598 ⁵
2	4,36	9,37	19,37	49,37	99,37	999	100 ²
3	2,98	5,25	8,85	17,01	27,49	131	609
4	2,47	3,95	6,04	10,16	14,80	49,00	157
5	2,20	3,34	4,82	7,50	10,29	27,65	71,23
6	2,04	2,98	4,15	6,14	8,10	19,03	42,54
7	1,93	2,75	3,73	5,33	6,84	14,63	29,64
8	1,86	2,59	3,44	4,79	6,03	12,05	22,71
9	1,80	2,47	3,23	4,41	5,47	10,37	18,50
10	1,75	2,38	3,07	4,13	5,06	9,20	15,74
11	1,72	2,30	2,95	3,91	4,74	8,35	13,80
12	1,69	2,24	2,85	3,74	4,50	7,71	12,38
13	1,66	2,20	2,77	3,60	4,30	7,21	11,30
14	1,64	2,15	2,70	3,48	4,14	6,80	10,46
15	1,62	2,12	2,64	3,39	4,00	6,47	9,78
16	1,61	2,09	2,59	3,30	3,89	6,19	9,23
17	1,59	2,06	2,55	3,23	3,79	5,96	8,76
18	1,58	2,04	2,51	3,17	3,71	5,76	8,38
19	1,57	2,02	2,48	3,12	3,63	5,59	8,04
20	1,56	2,00	2,45	3,07	3,56	5,44	7,76
21	1,55	1,98	2,42	3,02	3,51	5,31	7,51
22	1,54	1,97	2,40	2,99	3,45	5,19	7,29
23	1,53	1,95	2,37	2,95	3,41	5,09	7,09
24	1,53	1,94	2,36	2,92	3,36	4,99	6,92
25	1,52	1,93	2,34	2,89	3,32	4,91	6,76
26	1,52	1,92	2,32	2,86	3,29	4,83	6,62
27	1,51	1,91	2,31	2,84	3,26	4,76	6,50
28	1,51	1,90	2,29	2,82	3,23	4,69	6,38
29	1,50	1,89	2,28	2,80	3,20	4,64	6,28
30	1,50	1,88	2,27	2,78	3,17	4,58	6,18
40	1,47	1,83	2,18	2,64	2,99	4,21	5,53
60	1,44	1,77	2,10	2,51	2,82	3,86	4,95
100	1,41	1,73	2,03	2,41	2,69	3,61	4,53
140	1,40	1,71	2,01	2,37	2,64	3,51	4,36
∞	1,38	1,67	1,94	2,27	2,51	3,27	3,98

TABLA 10 Valores críticos de la distribución *F* (continuación)

gl del numerador = 9

gl del denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	14,68	59,86	241	151 ¹	602 ¹	602 ³	602 ⁵
2	4,37	9,38	19,38	49,39	99,39	999	100 ²
3	2,98	5,24	8,81	16,93	27,35	130	606
4	2,46	3,94	6,00	10,07	14,66	48,47	155
5	2,20	3,32	4,77	7,42	10,16	27,24	70,13
6	2,03	2,96	4,10	6,05	7,98	18,69	41,73
7	1,93	2,72	3,68	5,24	6,72	14,33	28,99
8	1,85	2,56	3,39	4,70	5,91	11,77	22,14
9	1,79	2,44	3,18	4,33	5,35	10,11	18,00
10	1,74	2,35	3,02	4,04	4,94	8,96	15,27
11	1,70	2,27	2,90	3,83	4,63	8,12	13,37
12	1,67	2,21	2,80	3,66	4,39	7,48	11,98
13	1,65	2,16	2,71	3,52	4,19	6,98	10,92
14	1,63	2,12	2,65	3,40	4,03	6,58	10,09
15	1,61	2,09	2,59	3,30	3,89	6,26	9,42
16	1,59	2,06	2,54	3,22	3,78	5,98	8,88
17	1,58	2,03	2,49	3,15	3,68	5,75	8,43
18	1,56	2,00	2,46	3,09	3,60	5,56	8,05
19	1,55	1,98	2,42	3,03	3,52	5,39	7,72
20	1,54	1,96	2,39	2,98	3,46	5,24	7,44
21	1,53	1,95	2,37	2,94	3,40	5,11	7,19
22	1,53	1,93	2,34	2,90	3,35	4,99	6,98
23	1,52	1,92	2,32	2,87	3,30	4,89	6,79
24	1,51	1,91	2,30	2,83	3,26	4,80	6,62
25	1,51	1,89	2,28	2,81	3,22	4,71	6,47
26	1,50	1,88	2,27	2,78	3,18	4,64	6,33
27	1,49	1,87	2,25	2,76	3,15	4,57	6,21
28	1,49	1,87	2,24	2,73	3,12	4,50	6,09
29	1,49	1,86	2,22	2,71	3,09	4,45	5,99
30	1,48	1,85	2,21	2,69	3,07	4,39	5,90
40	1,45	1,79	2,12	2,56	2,89	4,02	5,26
60	1,42	1,74	2,04	2,43	2,72	3,69	4,69
100	1,40	1,69	1,97	2,33	2,59	3,44	4,29
140	1,39	1,68	1,95	2,29	2,54	3,34	4,12
∞	1,36	1,63	1,88	2,19	2,41	3,10	3,75

TABLA 10 Valores críticos de la distribución *F* (continuación)

gl del numerador = 10

gl del denom.	Probabilidad de la cola						
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001
1	14,77	60,19	242	151 ¹	606 ¹	606 ³	606 ⁵
2	4,38	9,39	19,40	49,40	99,40	999	100 ²
3	2,98	5,23	8,79	16,86	27,23	129	603
4	2,46	3,92	5,96	10,00	14,55	48,05	154
5	2,19	3,30	4,74	7,34	10,05	26,92	69,25
6	2,03	2,94	4,06	5,98	7,87	18,41	41,08
7	1,92	2,70	3,64	5,17	6,62	14,08	28,45
8	1,84	2,54	3,35	4,63	5,81	11,54	21,68
9	1,78	2,42	3,14	4,26	5,26	9,89	17,59
10	1,73	2,32	2,98	3,97	4,85	8,75	14,90
11	1,69	2,25	2,85	3,76	4,54	7,92	13,02
12	1,66	2,19	2,75	3,59	4,30	7,29	11,65
13	1,64	2,14	2,67	3,45	4,10	6,80	10,60
14	1,62	2,10	2,60	3,33	3,94	6,40	9,79
15	1,60	2,06	2,54	3,23	3,80	6,08	9,13
16	1,58	2,03	2,49	3,15	3,69	5,81	8,60
17	1,57	2,00	2,45	3,08	3,59	5,58	8,15
18	1,55	1,98	2,41	3,02	3,51	5,39	7,78
19	1,54	1,96	2,38	2,96	3,43	5,22	7,46
20	1,53	1,94	2,35	2,91	3,37	5,08	7,18
21	1,52	1,92	2,32	2,87	3,31	4,95	6,94
22	1,51	1,90	2,30	2,83	3,26	4,83	6,73
23	1,51	1,89	2,27	2,80	3,21	4,73	6,54
24	1,50	1,88	2,25	2,77	3,17	4,64	6,37
25	1,49	1,87	2,24	2,74	3,13	4,56	6,23
26	1,49	1,86	2,22	2,71	3,09	4,48	6,09
27	1,48	1,85	2,20	2,69	3,06	4,41	5,97
28	1,48	1,84	2,19	2,66	3,03	4,35	5,86
29	1,47	1,83	2,18	2,64	3,00	4,29	5,76
30	1,47	1,82	2,16	2,62	2,98	4,24	5,66
40	1,44	1,76	2,08	2,49	2,80	3,87	5,04
60	1,41	1,71	1,99	2,36	2,63	3,54	4,48
100	1,38	1,66	1,93	2,26	2,50	3,30	4,08
140	1,37	1,64	1,90	2,22	2,45	3,20	3,93
∞	1,34	1,60	1,83	2,12	2,32	2,96	3,56

TABLA 11 Multiplicadores de Bonferroni para intervalos de confianza del 95 %Los valores dados en la tabla son $t_{gl, 0,025/k}$ siendo k el número de tests

gl	Número de tests									
	1	2	3	4	5	6	8	10	15	20
1	12,706	25,452	38,185	50,923	63,657	76,384	101,856	127,321	190,946	254,647
2	4,303	6,205	7,648	8,860	9,925	10,885	12,590	14,089	17,275	19,963
3	3,182	4,177	4,857	5,392	5,841	6,231	6,895	7,453	8,575	9,465
4	2,776	3,495	3,961	4,315	4,604	4,851	5,261	5,598	6,254	6,758
5	2,571	3,163	3,534	3,810	4,032	4,219	4,526	4,773	5,247	5,604
6	2,447	2,969	3,287	3,521	3,707	3,863	4,115	4,317	4,698	4,981
7	2,365	2,841	3,128	3,335	3,499	3,636	3,855	4,029	4,355	4,595
8	2,306	2,752	3,016	3,206	3,355	3,479	3,677	3,833	4,122	4,334
9	2,262	2,685	2,933	3,111	3,250	3,364	3,547	3,690	3,954	4,146
10	2,228	2,634	2,870	3,038	3,169	3,277	3,448	3,581	3,827	4,005
11	2,201	2,593	2,820	2,981	3,106	3,208	3,370	3,497	3,728	3,895
12	2,179	2,560	2,779	2,934	3,055	3,153	3,308	3,428	3,649	3,807
13	2,160	2,533	2,746	2,896	3,012	3,107	3,256	3,372	3,584	3,735
14	2,145	2,510	2,718	2,864	2,977	3,069	3,214	3,326	3,529	3,675
15	2,131	2,490	2,694	2,837	2,947	3,036	3,177	3,286	3,484	3,624
16	2,120	2,473	2,673	2,813	2,921	3,008	3,146	3,252	3,444	3,581
17	2,110	2,458	2,655	2,793	2,898	2,984	3,119	3,222	3,410	3,543
18	2,101	2,445	2,639	2,775	2,878	2,963	3,095	3,197	3,380	3,510
19	2,093	2,433	2,625	2,759	2,861	2,944	3,074	3,174	3,354	3,481
20	2,086	2,423	2,613	2,744	2,845	2,927	3,055	3,153	3,331	3,455
25	2,060	2,385	2,566	2,692	2,787	2,865	2,986	3,078	3,244	3,361
30	2,042	2,360	2,536	2,657	2,750	2,825	2,941	3,030	3,189	3,300
40	2,021	2,329	2,499	2,616	2,704	2,776	2,887	2,971	3,122	3,227
50	2,009	2,311	2,477	2,591	2,678	2,747	2,855	2,937	3,083	3,184
60	2,000	2,299	2,463	2,575	2,660	2,729	2,834	2,915	3,057	3,156
70	1,994	2,291	2,453	2,564	2,648	2,715	2,820	2,899	3,039	3,137
80	1,990	2,284	2,445	2,555	2,639	2,705	2,809	2,887	3,026	3,122
100	1,984	2,276	2,435	2,544	2,626	2,692	2,793	2,871	3,007	3,102
140	1,977	2,266	2,423	2,530	2,611	2,676	2,776	2,852	2,986	3,079
1.000	1,962	2,245	2,398	2,502	2,581	2,643	2,740	2,813	2,942	3,031
∞	1,960	2,241	2,394	2,498	2,576	2,638	2,734	2,807	2,935	3,023

SOLUCIONES DE EJERCICIOS SELECCIONADOS

Capítulo 1

1.2.3 El acupuntor espera que la acupuntura funcione mejor que la aspirina, de modo que es propenso a «ver» una mejora mayor en alguien que ha recibido acupuntura que en alguien que ha recibido aspirina (incluso aunque los dos grupos sean en verdad equivalentes entre sí en su respuesta al tratamiento).

1.3.1 (a) Muestreo por grupos. Las tres clínicas son los tres grupos.

1.3.2 (a) La muestra es no aleatoria y probablemente no representativa de la población general, porque consiste en (1) voluntarios de (2) clubes nocturnos. (i) El nivel de ansiedad social de la gente que asiste a clubes nocturnos es probablemente menor que el nivel de ansiedad social del público en general. (ii) Una estrategia de muestreo mejor sería escoger sujetos de toda la población.

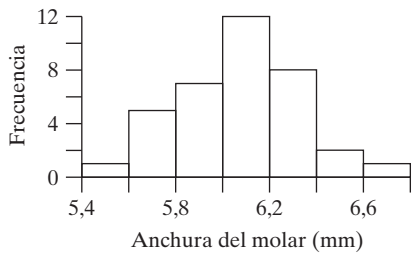
Capítulo 2

2.1.2 (a) (i) La altura y el peso. (ii) Variables continuas. (iii) Un niño. (iv) 37.

(b) (i) Grupo sanguíneo y nivel de colesterol. (ii) El grupo sanguíneo es categórica, el nivel de colesterol es continua. (iii) Una persona. (iv) 129.

2.2.1 (a) No hay una respuesta correcta única. Una posibilidad es

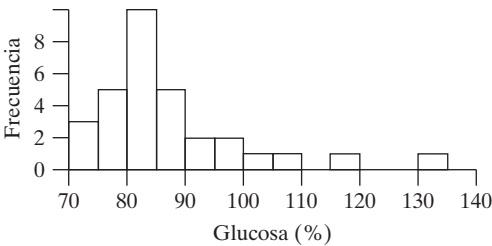
Anchura del molar	Frecuencia (n.º de especímenes)
[5,4, 5,6)	1
[5,6, 5,8)	5
[5,8, 6,0)	7
[6,0, 6,2)	12
[6,2, 6,4)	8
[6,4, 6,6)	2
[6,6, 6,8)	1
Total	36



(b) La distribución es bastante simétrica.

2.2.7 No hay una respuesta correcta única. Una posibilidad es

Glucosa (%)	Frecuencia (n.º de perros)
[70, 75)	3
[75, 80)	5
[80, 85)	10
[85, 90)	5
[90, 95)	2
[95, 100)	2
[100, 105)	1
[105, 110)	1
[110, 115)	0
[115, 120)	1
[120, 125)	0
[125, 130)	0
[130, 135)	1
Total	31



2.3.1 Cualquier muestra que cumpla $\sum y_i = 100$ sería una respuesta correcta; por ejemplo, 18, 19, 20, 21, 22.

2.3.5 $\bar{y} = 293,8$ mg/dl; mediana = 283 mg/dl.

2.3.6 $\bar{y} = 309$ mg/dl; mediana = 292 mg/dl.

2.3.11 Mediana = 10,5 cochinitos.

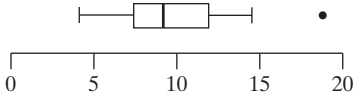
2.3.13 Media $\approx$ mediana $\approx$ 50.

2.3.14 15 %.

2.4.2 (a) Mediana = 9,2, $Q_1 = 7,4$, $Q_3 = 11,9$

(b) IQR = 4,5. **(c)** Valla superior = 18,65.

(d)



2.6.1 (a) $s = 2,45$. **(b)** $s = 3,32$.

2.6.4 (a) $\bar{y} = 33,10$ lb; $s = 3,444$ lb.

(b) Coef. de var. = 10,4 %.

2.6.9 (a) $32,23 \pm 8,07$ contiene 10/15 o el 67 % de las observaciones.

(b) 16,09 a 48,37 contiene 15/15 o el 100 % de las observaciones.

2.6.14 4 %.

2.6.15 $\bar{y} = 45$; $s = 12$.

2.7.1 Media = 37,3; DT = 12,9.

2.S.13 (a) Mediana = 38. **(b)** $Q_1 = 36$, $Q_3 = 41$.

(d) 66,4 %.

Capítulo 3

3.2.1 (a) 0,51. **(b)** 0,94. **(c)** 0,46. **(d)** 0,54.

3.2.5 (a) 0,107.

3.2.6 (a) 0,916.

3.3.1 (a) 0,185. **(b)** 0,117.

(c) No; $\Pr\{\text{Fumar}\} \neq \Pr\{\text{Fumar}|\text{Ingresos altos}\}$.

3.4.3 (a) 0,62.

3.5.5 0,9.

3.5.6 0,794.

3.6.6 (a) 0,3746. **(b)** 0,0688. **(c)** 0,1254.

3.6.9 (a) $0,75^6 = 0,1780$. **(b)** $1 - 0,1780 = 0,8220$.

3.7.1 Frecuencias esperadas: 939,5; 5.982,5; 15.873,1; 22.461,8; 17.879,3; 7.590,2; 1.342,6.

3.S.3 0,3369.

3.S.7 (a) $1 - 0,99^{100} = 0,6340$

(b) $1 - 0,99^n \geq 0,95$, por tanto $n \geq \log(0,05)/\log(0,99)$, por tanto, $n \geq 299$.

3.S.10 (a) 0,0209.

Capítulo 4

4.3.3 (a) 84,13 %. **(b)** 61,47 %. **(c)** 77,34 %.

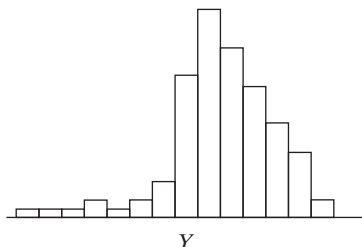
(d) 22,66 %. **(e)** 20,38 %. **(f)** 20,38 %.

4.3.4 (a) 22,66 %. **(b)** 20,38 %.

4.3.8 (a) 90,7 lb. **(b)** 85,3 lb.

4.3.12 (a) 98,76 %. **(b)** 98,76 %. **(c)** 1,24 %.

4.4.3. (b)



4.4.8 (a) No, el P valor pequeño proporciona evidencia de no normalidad.

4.S.4 (a) 97,98 %. **(b)** 12,71 %. **(c)** 46,39 %.

(d) 10,69 %. **(e)** 35,51 %. **(f)** 5,59 %. **(g)** 59,10 %.

4.S.5 0,122.

4.S.6 173,2 cm.

4.S.8 (a) 1,96. **(b)** 2,58.

4.S.15 0,1056.

4.S.18 200.

Capítulo 5

5.1.2 23/64

5.2.5 (a) 28,12 %. **(b)** 71,54 %. **(c)** 0,7154.

5.2.6 (a) 0,6680.

5.2.10 (a) 0,1861. **(b)** 0,9044.

5.2.13 (a) 0,1056. **(b)** 0,0150.

5.2.15 (a) 41,5. **(b)** 2,35.

5.3.1 (a) 0,66. **(b)** 0,29.

5.4.1 (a) 0,1762. **(b)** 0,1742.

5.4.4 (a) 0,7198.

5.4.6 (a) 0,9708.

5.S.1 (a) 0,2611.

5.S.9 9,68 %.

Capítulo 6

6.2.1 (a) 51,3 ng/g. **(b)** 26,5 ng/g.

6.3.3 (a) ET = 3,9 mg. **(b)** (23,4, 40,0).

6.3.10 (a) $4,1 < \mu < 21,9$ pg/ml.

(b) Tenemos una confianza del 95 % en que los niveles medios de BEH de enero a mayo en la población de todos los participantes en programas de educación física como el del estudio esté entre 4,1 y 21,9 pg/ml.

6.3.14 $1,17 < \mu < 1,23$ mm.

6.3.18 2,81.

6.4.2 178 hombres.

6.4.5 36 plantas.

6.5.1 La DT es mayor que la media, pero no son posibles valores negativos. Por tanto, la distribución debe estar sesgada hacia la derecha.

6.5.6 Las unidades observacionales (alícuotas) están anidadas dentro de las unidades (matraces), ubicadas aleatoriamente dentro de los tratamientos. Por tanto, las seis observaciones pueden no ser independientes entre sí debido a la estructura jerárquica de los datos, por lo que el método de cálculo del ET del investigador puede no ser válido.

6.6.1 2,41.

6.6.7 0,44.

6.7.1. $1,46 < \mu_1 - \mu_2 < 3,34$.

6.7.6. (a) $-5 < \mu_1 - \mu_2 < 9$ seg.

(c) Tenemos una confianza del 90 % en que el tiempo medio poblacional de la protrombina para ratas tratadas con un antibiótico (μ_1) es menor que el de las ratas de control (μ_2) en una cantidad que puede ser hasta de 5 segundos o que puede ser mayor que el de las ratas de control (μ_2) en una cantidad que puede ser hasta de 9 segundos.

6.S.2 (a) $\bar{y} = 2,275$; $s = 0,238$; $ET = 0,084$.

(b) (2,08, 2,47).

(c) μ = diámetro medio poblacional del tallo de plantas de trigo *Tetratichon* tres semanas después de florecer.

6.S.4 63 plantas.

6.S.9 (a) Debemos poder ver los datos como una muestra aleatoria de observaciones independientes de una población grande aproximadamente normal.

(b) Normalidad de la población.

(c) La independencia de las observaciones sería cuestionable, porque los pesos al nacer de los miembros de una pareja de gemelos podrían ser dependientes.

Capítulo 7

Observación sobre los contrastes de hipótesis La respuesta a un ejercicio sobre contraste de hipótesis incluye planteamientos verbales de las hipótesis y una afirmación verbal de la conclusión del contraste en el contexto del problema. Al expresar esas afirmaciones, hemos intentado capturar la esencia de la cuestión biológica que se considera. No obstante, las afirmaciones están necesariamente simplificadas y pasan por alto muchos aspectos que en la realidad podrían ser importantes. Por ejemplo, las hipótesis y la conclusión se podrían referir a una relación causal entre un tratamiento y una respuesta. En realidad, la validez de esa interpretación causal depende generalmente de varios factores relacionados con el diseño de la investigación (como una asignación no sesgada de animales a grupos de tratamiento) y con los procedimientos experimentales concretos (como la precisión de los ensayos o de las técnicas de medida). En resumen, el estudiante debe advertir que las afirmaciones verbales intentan clarificar los conceptos *estadísticos*; su contenido *biológico* puede ser una cuestión abierta.

7.1.2 (b) 9.

7.2.1. (a) $t_s = -3,13$, por tanto $0,02 < P \text{ valor} < 0,04$.

(b) $t_s = 1,25$, por tanto $0,20 < P < 0,40$.

(c) $t_s = 4,62$, por tanto $P < 0,001$.

7.2.3. (a) Sí. **(b)** No. **(c)** Sí. **(d)** No.

7.2.7 (a) H_0 : la concentración media de serotonina es la misma en los pacientes de corazón que en los controles ($\mu_1 = \mu_2$); H_A : la concentración media de serotonina no es la misma en los pacientes de corazón que en los controles ($\mu_1 \neq \mu_2$). $t_s = -1,38$. No se rechaza H_0 . No hay suficiente evidencia ($0,10 < P < 0,20$) para concluir que los niveles de

serotonina son diferentes en los pacientes de corazón y en los controles.

7.2.11 H_0 : la inundación no tiene efecto en el ATP ($\mu_1 = \mu_2$); H_A : la inundación tiene algún efecto en el ATP ($\mu_1 \neq \mu_2$). $t_s = -3,92$. Se rechaza H_0 .

7.3.4 Tipo II.

7.3.6 Sí; dado que el cero está fuera del intervalo de confianza, sabemos que el P valor es menor que 0,05, por lo que rechazamos la hipótesis de que $\mu_1 - \mu_2 = 0$.

7.4.1 Las personas con problemas respiratorios se trasladan a Arizona (porque el aire seco es bueno para ellas).

7.4.4 (a) El número de tazas de café.

(b) Enfermedad de la arteria coronaria (presente o ausente).

(c) Los sujetos (es decir, las 1.040 personas).

7.5.1 (a) $0,10 < P < 0,20$. **(b)** $0,03 < P < 0,04$.

7.5.3 (a) Sí. **(b)** Sí. **(c)** Sí. **(d)** No.

7.5.9 H_0 : dañar la planta no tiene efecto en el crecimiento de las larvas ($\mu_1 = \mu_2$). H_A : dañar la planta tiende a disminuir el crecimiento de las larvas ($\mu_1 < \mu_2$), donde 1 indica las plantas dañadas y 2 las de control. $t_s = -2,69$. Se rechaza H_0 . Hay suficiente evidencia ($0,005 < P < 0,01$) para concluir que dañar las plantas tiende a disminuir el crecimiento de las larvas.

7.5.10 (a) H_0 : el medicamento no tiene efecto sobre el dolor ($\mu_1 = \mu_2$); H_A : el medicamento aumenta el alivio del dolor ($\mu_1 > \mu_2$). $t_s = 1,81$. Se rechaza H_0 . Hay suficiente evidencia ($0,03 < P < 0,04$) para concluir que el medicamento es efectivo.

(b) El P valor estaría entre 0,06 y 0,08. Con $\alpha = 0,05$ no rechazaríamos H_0 .

7.6.4 No, de acuerdo al intervalo de confianza los datos no indican si la verdadera diferencia es «importante».

7.6.6 0,33.

7.7.1 (a) 23. **(b)** 11.

7.7.4 (a) 71. **(b)** 101. **(c)** 58.

7.7.6 0,5.

7.10.1 (a) $P > 0,149$. **(b)** $P = 0,048$. **(c)** $P = 0,0025$.

7.10.3 (a) H_0 : el tolueno no tiene efecto sobre la dopamina en el striatum de las ratas; H_A : el tolueno tiene algún efecto sobre la dopamina en el striatum de las ratas. $U_s = 32$. Se rechaza H_0 . Hay suficiente evidencia ($0,015 < P < 0,041$) para concluir que el tolueno aumenta la dopamina en el striatum de las ratas.

7.S.2 H_0 : el calcio medio en las plaquetas es el mismo en personas con la presión sanguínea alta que en las personas con la presión sanguínea normal ($\mu_1 = \mu_2$); H_A : el calcio medio en las plaquetas es diferente en las personas con la presión sanguínea alta que en las personas con la presión sanguínea normal ($\mu_1 \neq \mu_2$). $t_s = 11,2$. Se rechaza H_0 . Hay suficiente evidencia ($P < 0,0001$) para concluir que el calcio

en las plaquetas es mayor en personas con la presión sanguínea alta.

7.S.4 No; el test t es válido porque los tamaños muestrales son bastante grandes.

7.S.8 H_0 : la tensión no afecta al crecimiento; H_A : la tensión tiende a retrasar el crecimiento. $U_s = 148,5$. Se rechaza H_0 . Hay suficiente evidencia ($P < 0,0021$) para concluir que la tensión tiende a retrasar el crecimiento.

7.S.21 Falso: el cero está en el intervalo de confianza.

Capítulo 8

8.2.1 (a) 0,34.

8.2.3 H_0 : la progesterona no tiene efecto sobre el cAMP ($\mu_1 = \mu_2$); H_A : la progesterona tiene algún efecto sobre el cAMP ($\mu_1 \neq \mu_2$). $t_s = 3,4$. Se rechaza H_0 . Hay suficiente evidencia ($0,04 < P \text{ valor} < 0,05$) para concluir que la progesterona disminuye el cAMP bajo estas condiciones.

8.2.6 (a) $-0,50 < \mu_1 - \mu_2 < 0,74$ °C, donde 1 indica tratamiento y 2 indica control.

8.4.1 (a) $P > 0,20$. **(b)** $P = 0,180$. **(c)** $P = 0,039$. **(d)** $P = 0,004$.

8.4.4 H_0 : el entorno no afecta al peso del córtex cerebral ($p = 0,5$); H_A : el entorno enriquecido aumenta el peso del córtex ($p > 0,5$). $B_s = 10$. No se rechaza H_0 . No hay suficiente evidencia ($P = 0,0195$) para concluir que el entorno enriquecido aumenta el peso del córtex.

8.4.8 0,000061.

8.4.11 $n = 6$; $P \text{ valor} = 0,03125$.

8.5.1 (a) $P > 0,20$. **(b)** $P = 0,078$. **(c)** $P = 0,047$. **(d)** $P = 0,016$.

8.5.3 H_0 : el tratamiento (mCFP versus placebo) no afecta la valoración del hambre; H_A : el tratamiento afecta a la valoración del hambre. $W_s = 27$ y $n_D = 8$. Se rechaza H_0 . Hay suficiente evidencia ($P > 0,20$) para concluir que el tratamiento tiene efecto.

8.6.4 No. Predicción «exacta» significaría que las diferencias individuales (d) son pequeñas. Para juzgar si este es el caso, serían necesarios los valores individuales de las d . Con ellos, se podría ver si los módulos ($|d|$) son pequeños.

8.S.8 H_0 : el número medio de especies es el mismo en los remansos y en los rápidos ($\mu_1 = \mu_2$); H_A : el número medio de especies en los remansos y en los rápidos son diferentes ($\mu_1 \neq \mu_2$). $t_s = 4,58$. Se rechaza H_0 . Hay suficiente evidencia ($P < 0,001$) para concluir que el número medio de especies es mayor en los remansos que en los rápidos.

8.S.12 H_0 : la cafeína no tiene efecto sobre el RER ($\mu_1 = \mu_2$); H_A : la cafeína tiene algún efecto sobre el RER ($\mu_1 \neq \mu_2$). $t_s = 3,94$. Se rechaza H_0 . Hay suficiente evidencia ($0,001 < P < 0,01$) para concluir que la cafeína tiende a disminuir el RER bajo estas condiciones.

Capítulo 9

9.1.2 (a) 0,250. **(b)** 0,441.

(c) No; el mínimo número posible de mutantes es cero, en cuyo $\tilde{p}$ caso es $2/7$.

9.1.4 (a) 0,2501. **(b)** 0,0352.

9.1.5 (a) (i) 0,3164, (ii) 0,4219, (iii) 0,2109, (iv) 0,0469, (v) 0,0039.

9.1.9 0,5259.

9.2.2 (a) 0,040. **(b)** 0,020.

9.2.3 (a) (0,134, 0,290). **(b)** (0,164, 0,242).

9.2.5. (a) (0,164, 0,250).

(b) Tenemos una confianza del 95 % en que la probabilidad de una reacción adversa en niños que reciben su primera reacción de la vacuna esté entre 0,164 y 0,250.

9.2.7 $n \geq 146$.

9.3.4 (0,646, 0,838).

9.4.1 H_0 : la razón poblacional es 12:3:1 ($\Pr\{\text{blanco}\} = 0,75$, $\Pr\{\text{amarillo}\} = 0,1875$, $\Pr\{\text{verde}\} = 0,0625$); H_A : la razón no es 12:3:1. $\chi_s^2 = 0,69$. No se rechaza H_0 . Hay muy poca o ninguna evidencia ($P > 0,20$) a favor de que el modelo no es correcto. Los datos son consistentes con el modelo.

9.4.2 H_0 y H_A como en el Ejercicio 9.4.1. $\chi_s^2 = 6,9$. Se rechaza H_0 . Hay suficiente evidencia ($0,02 < P < 0,05$) para concluir que el modelo es incorrecto. Los datos no son consistentes con el modelo.

9.4.8 H_0 : la sustancia no causa tumores ($\Pr\{T\} = \frac{1}{3}$); H_A : la

sustancia causa tumores ($\Pr\{T\} > \frac{1}{3}$), siendo T el suceso de que un tumor se produzca primero en la rata tratada. $\chi_s^2 = 6,4$. Se rechaza H_0 . Hay suficiente evidencia ($0,005 < P < 0,01$) para concluir que la sustancia causa tumores.

9.S.2 (a) 0,2111. **(b)** 0,5700.

9.S.3 (0,707, 0,853).

9.S.14 (a) H_0 : la elección direccional es aleatoria ($\Pr\{\text{acercarse}\} = 0,25$, $\Pr\{\text{alejarse}\} = 0,25$, $\Pr\{\text{derecha}\} = 0,25$, $\Pr\{\text{izquierda}\} = 0,25$); H_A : la elección direccional no es aleatoria, siendo T el suceso de que un tumor se produzca primero en la rata tratada. $\chi_s^2 = 4,88$. No se rechaza H_0 . No hay suficiente evidencia ($0,10 < P < 0,20$) para concluir que la elección direccional no es aleatoria.

9.S.16 H_0 : la probabilidad de que un huevo esté sobre un tipo particular de judía es de 0,25 para los cuatro tipos de judías; H_A : H_0 es falsa. $\chi_s^2 = 2,23$. No se rechaza H_0 . No hay suficiente evidencia ($P > 0,20$) para concluir que los gorgojos de la caupí prefieren un tipo de judía sobre las demás.

Capítulo 10

10.2.3 (a)

5	20
10	40

(b) $\hat{p}_1 = 5/15 = 1/3$ y $\hat{p}_2 = 20/60 = 1/3$; sí.

10.2.5 H_0 : los ácaros no inducen resistencia al hongo ($p_1 = p_2$); H_A : los ácaros inducen resistencia al hongo ($p_1 < p_2$), siendo p la probabilidad de contraer enfermedad por hongo y 1 indica ácaros y 2 no ácaros. $\chi_s^2 = 7,21$. Se rechaza H_0 . Hay suficiente evidencia ($0,0005 < P$ valor $< 0,005$) para concluir que los ácaros inducen resistencia al hongo.

10.2.10 H_0 : las dos formas de administración son igual de efectivas ($p_1 = p_2$); H_A : las dos formas de administración no son igual de efectivas ($p_1 \neq p_2$). $\chi_s^2 = 4,48$. Se rechaza H_0 . Hay suficiente evidencia ($0,02 < P$ valor $< 0,05$) para concluir que la administración simultánea es superior a la administración secuencial.

10.2.13 H_0 : el ancrod y el placebo son igual de efectivos ($p_1 = p_2$); H_A : el ancrod y el placebo no son igual de efectivos ($p_1 \neq p_2$). $\chi_s^2 = 3,82$. No se rechaza H_0 . No hay suficiente evidencia ($0,05 < P$ valor $< 0,10$) para concluir que los tratamientos se diferencian.

10.3.3 (a) $\hat{\Pr}\{F|C\} = 0,239$, $\hat{\Pr}\{F|EV\} = 0,305$, $\hat{\Pr}\{C|F\} = 0,439$, $\hat{\Pr}\{D|V\} = 0,522$.

(b) H_0 : no hay asociación entre tratamiento y supervivencia ($\Pr\{F|C\} = \Pr\{F|EV\}$); H_A : existe alguna asociación entre el método de tratamiento (cirugía frente a espera vigilante) y la supervivencia ($\Pr\{F|C\} \neq \Pr\{F|EV\}$). No se rechaza H_0 . No hay suficiente evidencia ($0,05 < P$ valor $< 0,10$) para concluir que las tasas de supervivencia son diferentes para los dos tratamientos.

10.3.4 (a) $\hat{\Pr}\{PD|MD\} = 0,934$. (b) $\hat{\Pr}\{PD|MI\} = 0,511$.
(c) $\chi_s^2 = 398$. (d) $\chi_s^2 = 1,623$.

10.4.1

5	1
9	15

6	0
8	16

10.5.3 (a) H_0 : la distribución de los grupos sanguíneos son las mismas para los pacientes con úlcera que para los de control ($\Pr\{O|PU\} = \Pr\{O|C\}$, $\Pr\{Q|PU\} = \Pr\{A|C\}$, $\Pr\{B|PU\} = \Pr\{B|C\}$, $\Pr\{AB|PU\} = \Pr\{AB|C\}$); H_A : la distribución de los grupos sanguíneos no son las mismas. Se rechaza H_0 . Hay suficiente evidencia (P valor $< 0,0001$,

$gl = 3$) para concluir que la distribución de los grupos sanguíneos de los pacientes con úlcera es diferente de la de los pacientes de control.

10.5.5 (a) H_0 : el cambio en la escala EEEA-Cog es independiente del tratamiento; H_A : el cambio en la escala EEEA-Cog está relacionado con el tratamiento. $\chi_s^2 = 10,26$, $gl = 4$. Se rechaza H_0 . Hay suficiente evidencia ($0,02 < P$ valor $< 0,05$) para concluir que el EGB y el placebo no son igual de efectivos.

10.6.2 Este análisis no es apropiado porque las unidades observacionales (ratones) están anidadas dentro de las unidades (camadas) asignadas aleatoriamente a los tratamientos. Esta estructura jerárquica arroja dudas sobre la condición de que las observaciones de los 224 ratones son independientes, especialmente a la luz del comentario del investigador indicando que la respuesta varía considerablemente de una camada a otra.

10.7.3 $0,001 < p_1 - p_2 < 0,230$. No; el intervalo de confianza sugiere que el reposo en cama podría resultar de hecho dañino.

10.7.5 (a) $0,067 < p_1 - p_2 < 0,118$.

(b) Tenemos una confianza del 95% en que la proporción de personas con sangre de grupo O entre los pacientes de úlcera es mayor que la proporción de pacientes con sangre de tipo O entre individuos sanos en un valor que puede estar entre 0,067 y 0,118. Es decir, tenemos una confianza del 95 % en que p_1 supera a p_2 por un valor que está entre 0,067 y 0,118.

10.8.1 H_0 : no hay asociación entre el uso de anticonceptivos orales y la hemorragia cerebral ($p = 0,5$); H_A : hay asociación entre el uso de anticonceptivos orales y la hemorragia cerebral ($p \neq 0,5$), donde p indica la probabilidad de que una pareja discordante sea Sí(caso)/No(control). $\chi_s^2 = 6,72$. Se rechaza H_0 . Hay suficiente evidencia ($0,001 < P$ valor $< 0,01$) para concluir que es más probable que las víctimas de hemorragia cerebral utilicen anticonceptivos orales ($p > 0,5$).

10.9.1 (a) (i) 1,339, (ii) 1,356. (b) (i) 1,314, (ii) 1,355.

10.9.7 (a) 1,241. (b) (1,036, 1,488).

(c) Tenemos una confianza del 95 % en que la heparina aumenta la ventaja de una respuesta negativa en un factor que está entre 1,036 y 1,488, en comparación con tomar enoxaparina.

10.S.3 (a) H_0 : la razón de sexos es 1:1 en un entorno caliente ($p_1 = 0,5$); H_A : la razón de sexos no es 1:1 en un entorno caliente ($p_1 \neq 0,5$), donde p_1 indica la probabilidad de una hembra en un entorno caliente. $\chi_s^2 = 0,18$. No se rechaza H_0 . No hay suficiente evidencia (P valor $> 0,20$) para concluir que la razón de sexos no es 1:1 en un entorno caliente.

(c) H_0 : la razón de sexos es la misma en los dos entornos ($p_1 = p_2$); H_A : la razón de sexos no es la misma en los dos entornos ($p_1 \neq p_2$), siendo p la probabilidad de una hembra y

1 y 2 indican los entornos caliente y frío. $\chi^2_s = 4,20$. Se rechaza H_0 . Hay suficiente evidencia ($0,02 < P \text{ valor} < 0,05$) para concluir que la probabilidad de una hembra es mayor en el entorno frío que en el caliente.

10.S.12 H_0 : los sitios de captura y de recaptura son independientes ($P\{RI|CI\} = P\{RI|CII\}$); H_A : las moscas vuelven preferentemente a su lugar de captura ($P\{RI|CI\} > P\{RI|CII\}$), donde C y R indican captura y recaptura y I y II indican los lugares. Se rechaza H_0 . Hay suficiente evidencia ($0,0005 < P \text{ valor} < 0,005$) para concluir que las moscas vuelven preferentemente a su lugar de captura.

10.S.14 (a) 1,709. **(b)** $1,55 < \theta < 1,89$.

(c) La razón de ventajas da la ventaja (estimada) de supervivencia de los hombres comparados con las mujeres. Esta razón (de 1,709) es una buena aproximación al riesgo relativo de fallecimiento de mujeres comparadas con hombres (que es 1,658), debido a que el fallecimiento es bastante raro.

Capítulo 11

11.2.1 (a) $SC(\text{entre}) = 228$, $SC(\text{intra}) = 120$.

(b) $SC(\text{total}) = 348$.

(c) $CM(\text{entre}) = 114$, $CM(\text{intra}) = 15$, $s_{\text{conjunta}} = 3,87$.

11.2.4 (a)

Fuente	gl	SC	CM
Entre grupos	3	135	45
Dentro de los grupos	12	337	28,08
Total	15	472	

(b) 4. **(c)** 16.

11.4.2 (a) H_0 : las condiciones de estrés producen la misma concentración media de linfocitos ($\mu_1 = \mu_2 = \mu_3 = \mu_4$); H_A : algunas condiciones de stress producen diferentes concentraciones medias de linfocitos (las μ no son iguales). $F_s = 3,84$. Se rechaza H_0 . Hay suficiente evidencia ($0,01 < P \text{ valor} < 0,02$) para concluir que algunas condiciones de stress producen diferentes concentraciones medias de linfocitos.

(b) $s_{\text{conjunta}} = 2,78 \text{ células/ml} \times 10^{-6}$.

11.4.3 (a) H_0 : la BEH media es la misma en las tres poblaciones ($\mu_1 = \mu_2 = \mu_3$); H_A : la BEH media no es la misma en las tres poblaciones (las μ no son iguales). $F_s = 0,58$. No se rechaza H_0 . No hay suficiente evidencia ($P \text{ valor} > 0,20$) para concluir que la BEH media no es la misma en las tres poblaciones.

(b) $s_{\text{conjunta}} = 14,4 \text{ pg/ml}$.

11.6.2 No hay una única respuesta correcta. Una posibilidad es la siguiente:

Tratamiento	Cochinillo				
	Camada 1	Camada 2	Camada 3	Camada 4	Camada 5
1	2	5	2	4	5
2	1	4	1	1	2
3	4	2	5	2	4
4	5	3	3	3	3
5	3	1	4	5	1

11.6.5 El plan II es mejor. Deseamos que las unidades dentro de un bloque sean similares entre sí, y el plan II lo consigue. Bajo el plan I el efecto de la lluvia se podría confundir con el efecto de la variedad.

11.7.2 (a)

Fuente	gl	SC	CM
Entre especies	1	2,19781	2,19781
Entre niveles de inundación	1	2,25751	2,25751
Interacción	1	0,097656	0,097656
Dentro de los grupos	12	0,47438	0,03953
Total	15	5,027356	

(b) $F_s = 0,097656/0,03953 = 2,47$. Con $gl = 1$ y 12. En la Tabla 10, $F_{0,20} = 1,84$ y $F_{0,10} = 3,18$. Por tanto, $0,10 < P \text{ valor} < 0,20$ y no rechazamos H_0 . No hay suficiente evidencia ($P \text{ valor} > 0,10$) para concluir que hay una interacción presente.

(c) $F_s = 2,19781/0,03953 = 55,60$. Con $gl = 1$ y 12. En la Tabla 10, $F_{0,0001} = 32,43$. Por tanto, $P \text{ valor} < 0,0001$ y rechazamos H_0 . Hay suficiente evidencia ($P \text{ valor} < 0,0001$) para concluir que la especie afecta a la concentración de ATP.

(d) $s_{\text{conjunta}} = \sqrt{0,03953} = 0,199$.

11.7.4 (a) $F_s = \frac{31,33/1}{30,648,81/(223-4)} = 31,33/139,95 = 0,22$.

Con $gl = 1$ y 140. En la Tabla 10, $F_{0,20} = 1,66$. Por tanto, $P \text{ valor} > 0,20$ y no rechazamos H_0 . No hay suficiente evidencia ($P \text{ valor} > 0,20$) para concluir que hay interacción presente.

11.8.2 (a) 123 mmHg. **(b)** 123,2 mmHg.

(d) 0,851 mmHg.

11.8.7 $0,67 < \mu_E - \mu_S < 1,48 \text{ g}$, donde

$$\mu_E = \frac{1}{2}(\mu_{E, \text{bajo}} + \mu_{E, \text{alto}}) \text{ y } \mu_S = \frac{1}{2}(\mu_{S, \text{bajo}} + \mu_{S, \text{alto}}).$$

11.8.8 (b) $L = 3,685 \text{ nmol}/10^8 \text{ plaquetas/hora}$;
 $ET_L = 1,048 \text{ nmol}/10^8 \text{ plaquetas/hora}$.

11.9.1. Se rechazan las siguientes hipótesis: $H_0: \mu_C = \mu_D$; $H_0: \mu_A = \mu_D$; $H_0: \mu_B = \mu_D$; $H_0: \mu_C = \mu_E$; $H_0: \mu_A = \mu_E$; $H_0: \mu_B = \mu_E$; $H_0: \mu_B = \mu_C$; $H_0: \mu_A = \mu_C$. No se rechazan las siguientes hipótesis: $H_0: \mu_A = \mu_B$; $H_0: \mu_D = \mu_E$. Resumen:

C A B E D

Hay suficiente evidencia para concluir que los tratamientos D y E tienen las medias mayores, seguidas por los tratamientos A y B y el C la menor. No hay suficiente evidencia para concluir que los tratamientos A y B tienen medias diferentes ni que los tratamientos D y E tienen medias diferentes.

11.9.2 No se rechazan las siguientes hipótesis: $H_0: \mu_A = \mu_B$; $H_0: \mu_B = \mu_D$; $H_0: \mu_B = \mu_E$; $H_0: \mu_D = \mu_E$. Resumen:

C A B E D

11.9.4 (a) Sí, cada una de las dietas B, C y D es diferente de la A, ya que ninguno de los intervalos incluye al cero.

11.S.1 H_0 : las tres clases producen el mismo cambio medio en la masa libre de grasa ($\mu_1 = \mu_2 = \mu_3$); H_A : al menos una clase produce una media diferente (las μ no son iguales). $F_s = 0,64$. No se rechaza H_0 . No hay suficiente evidencia (P valor $> 0,20$) para concluir que las medias poblacionales son diferentes.

11.S.3 H_0 : el error refractivo medio es el mismo en las tres poblaciones ($\mu_1 = \mu_2 = \mu_3 = \mu_4$); H_A : algunas de las poblaciones tienen diferentes errores refractivos medios (las μ no son iguales). $F_s = 3,56$. Se rechaza H_0 . Hay suficiente evidencia ($0,01 < P$ valor $< 0,02$) para concluir que las poblaciones tienen diferentes errores refractivos medios.

11.S.13 Sean 1, 2, 3 y 4 el placebo, el probucol, las multivitaminas y el probucol con multivitaminas, respectivamente.

(a) $\bar{y}_2 - \bar{y}_1 = 1,79 - 1,43 = 0,36$.

(b) $\bar{y}_4 - \bar{y}_3 = 1,54 - 1,40 = 0,14$.

(c) El contraste que mide la interacción entre el probucol y las multivitaminas es «la diferencia de diferencias», de los apartados (a) y (b):

$$(\bar{y}_2 - \bar{y}_1) - (\bar{y}_4 - \bar{y}_3) = 0,36 - 0,14 = 0,22$$

(Nota: no es la única respuesta correcta; invertir los signos de (a) y (b), o de (c) es también correcto).

Capítulo 12

12.2.1 (d), (a), (b), (c), (e). (Las correlaciones son $-0,97$, $-0,63$, $0,10$, $0,58$ y $0,93$).

12.2.2 (b) $r = 0,439$.

12.2.3 H_0 : no hay correlación entre la urea en sangre y la concentración de ácido úrico ($\rho = 0$); H_A : la urea en sangre y la concentración de ácido úrico están correlados positivamente ($\rho > 0$). $t_s = 3,952$. Se rechaza H_0 . Hay suficiente evidencia (P valor $< 0,0005$) para concluir que la urea en sangre y la concentración de ácido úrico están correlados positivamente.

12.2.5 (a) H_0 : no hay correlación entre la densidad de plantas y el peso medio de la mazorca ($\rho = 0$); H_A : la densidad de plantas y el peso medio de la mazorca están correlacionados ($\rho \neq 0$). $t_s = -11,9$. Se rechaza H_0 . Hay suficiente evidencia (P valor $< 0,001$) para concluir que la densidad de plantas y el peso medio de la mazorca están correlados negativamente.

(b) Unidad observacional.

(c) No; se trata de un estudio observacional en el que la densidad de plantas se observa pero no se modifica. El estudio sugiere que puede merecer la pena modificar controladamente la densidad de plantas en un experimento posterior.

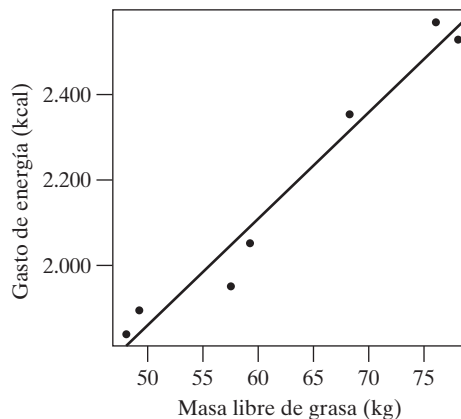
12.3.1 (b) Leucina $= -0,05 + 0,02928 \times \text{Tiempo}$; la pendiente es $0,02928$ ng/min.

(d) $s_e = 0,0839$.

12.3.2 (c) $\hat{y} = -0,592 + 7,641x$; $s_e = 0,881$ °C.

12.3.5 (a) $\hat{y} = 607,7 + 25,01x$.

(b)



(c) Cuando la masa libre de grasa aumenta 1 kg, el gasto de energía aumenta 25,01 kcal, en promedio.

(d) $s_e = 64,85$ kcal.

12.3.8 (b) $r^2 = 0,107 = 10,7$ %. (f) $12/17 = 71$ %.

12.4.5 Media estimada = 21,1 mm; DT estimada = 1,3 mm.

12.4.9 Media estimada = 658,1 l/min; DT estimada = 115,16 l/min.

12.5.1 (a) $0,0252 < \beta_1 < 0,0334$ ng/min.

(b) Tenemos un 95 % de confianza en que la velocidad con la que la leucina se incorpora a la proteína en la población de todos los ovocitos de *Xenopus* esté entre 0,0252 ng/min y 0,0334 ng/min.

12.5.5 (a) $19,4 < \beta_1 < 30,6$ kcal/kg.

(b) $20,6 < \beta_1 < 29,4$ kcal/kg.

12.5.7 H_0 : no hay relación lineal entre la tasa de respiración y la altitud en el origen ($\beta_1 = 0$); H_A : los árboles de altitudes mayores tienden a tener tasas de respiración más altas

($\beta_1 > 0$). $t_s = 6,06$. Se rechaza H_0 . Hay suficiente evidencia (P valor $< 0,0005$) para concluir que los árboles de altitudes mayores tienden a tener tasas de respiración más altas.

12.6.6 (a) – (iii). **(b)** – (i). **(c)** – (ii).

12.7.1 (a) Las líneas discontinuas, que nos dicen dónde están las verdaderas rectas de regresión (poblacionales).

12.S.1 (a) 0,24 g.

12.S.3 (a) Media estimada = 0,85 kg; DT estimada = 0,17 kg.

12.S.6 (a) $s_e = 0,137$ cm.

(b) $H_0: \rho = 0$ o $H_0: \beta_1 = 0$. $t_s = 3,01$. Se rechaza H_0 . Hay suficiente evidencia ($0,02 < P$ valor $< 0,04$) para concluir que existe una correlación positiva entre el diámetro de las ramas y la longitud de las alas.

Capítulo 13

13.2.1 Sería apropiado un test chi-cuadrado de independencia. La hipótesis nula de interés es $H_0: p_1 = p_2$, siendo $p_1 = \Pr\{\text{mejora clínica importante con clozapina}\}$ y $p_2 = \Pr\{\text{mejora clínica importante con haloperidol}\}$.

13.2.10 Aquí se requiere una correlación de dos muestras, pero los datos no permiten fundamentar la condición de normalidad. Por tanto, resulta apropiado usar el test de Wilcoxon-Mann-Whitney.

13.2.12 Sería natural considerar la correlación y la regresión con estos datos. Por ejemplo, podríamos calcular la regresión de $Y = \text{longitud del antebrazo}$ con respecto a $X = \text{altura}$; podríamos calcular también la correlación entre la longitud del antebrazo y la altura y contrastar la hipótesis nula de que la correlación poblacional es cero.

ÍNDICE

booksmedicos.org

A

Ácido giberélico (AG), 559
Alanina aminotransferasa (ALT), 37
Alternativa no direccional, 252
Análisis de correlación, 485
Análisis de covarianza, 540-541
Análisis de datos, 559. Véase Análisis exploratorio de datos
Análisis de varianza (ANOVA), 419
 «entre grupos», 430
 aplicabilidad de métodos, 437-440
 básico de un factor, 423
 cantidades con fórmulas, 430
 condiciones estándar, 437
 de dos factores, 452-459
 de un factor, 423
 dentro de los grupos, 430
 medida de la variación, 425
 desviación típica, 425
 efecto de los grupos, 432
 factorial, 452-458
 hipótesis nula, 431
 igualdad de DT de las poblaciones, 437-438
 modelo, 431
 notación, 425-426
 perspectiva gráfica, 422-423
 prueba F global, 471
 relación fundamental, 428
 tabla, 429
 verificación de condiciones, 437
Análisis exploratorio de datos, 559
Anécdota, 7, 184
ANOVA. Véase Análisis de varianza (ANOVA)
Aproximación mediante la normal a la distribución binomial, 164-168
Árbol de probabilidad, 89
Área de regiones de extensión infinita, 576-577

B

Bimodalidad, 34

C

Capacidad vital forzada (CVF), 459, 462
Características de forma, 35
Clases, 31
CM(bloques), 446
CM(intra), 426
CM(entre), 427
Codificación, 67
Coeficiente binomial, 109, 574-575
Coeficiente de correlación, 487, 489, 546
 causalidad, 493
 correlación y, 493
 reproducción de algas, 493
 fórmula, 489
 hipótesis nula, 491
 inferencia, 491
 interpretación, 490-491
 correlación muestral, 490
 correlación poblacional, 490
 intervalo de confianza:
 de la correlación poblacional, 492
 medida de fuerza de asociación lineal, 488
 modelo de muestreo aleatorio bivariado, 490-491
Coeficiente de determinación, 506-507
Coeficiente de variación, 61
Combinaciones lineales, 460
 contrastes, 460
 aberraciones cromosómicas, 465
 para evaluar interacciones, 463
 error típico (ET), 461
 intervalos de confianza, 462
 para ajuste, 460
 media «ajustada por edad», 460
 test t , 463
Comisura anterior (CA), 7
Comparación de medias de muestras independientes, 419

ANOVA, 420
 de dos factores, 452
 combinaciones lineales, 459
 comparaciones múltiples, 468
 diseño de bloques aleatorizados, 441, 444
 diseño de experimentos, 478
 eficiencia de tratamientos de métodos orgánicos, 419
 limitaciones del test t , 420
 estimación de la desviación típica, 421
 estructura en grupos, 421
 problema de las comparaciones múltiples, 421
 planteamientos no paramétricos, 478
 selección y ordenación, 478
 test F global, 433
 ventajas de un planteamiento global, 477
Comparaciones de medias, 201
 error típico (ET):
 de la diferencia de dos medias muestrales, 201
 experimento de tonsilectomía, 204
 error típico agrupado, 204
 cálculo de la capacidad vital, 205-206
 desviación típica (DT), 205
 estudios observacionales, 247-249
 notación, 201-202
Comparaciones de muestras pareadas
 análisis de diferencias, 301-303
 análisis de t -Student:
 condiciones de validez, 307
 fórmulas, 307
 ignorar el resultado del emparejamiento, 304
 intervalo de confianza, 303-304
 diagrama de puntos de las diferencias, 305
 diagrama de puntos paralelos, 306
 error típico (ET), 306

- Comparaciones múltiples, 468, 477
condiciones de validez, 475
DSH de Tukey, 474
DSM de Fisher, 468
error por experimento frente a error por comparación, 468
método de Bonferroni, 473
problema, 420
Confusión, 247-248
Contagio, 113
Contraste(s), 460
de bondad de ajuste, 352, 354
de hipótesis, 227
para evaluar la interacción, 463-465
Controles históricos, 12
Corrección de continuidad, 166-168
Correlación muestral, 490
Creatina fosfoquinasa (CFK), 31
Criterio de mínimos cuadrados, 504, 539
Cuadrados medios dentro de los grupos (CM(intra)), 426
Cuadrados medios entre grupos (CM(entre)), 427
Cuadrados medios por bloques (CM(bloques)), 446
Cuartiles, 44
Cuidado residencial de alto nivel (CRAN), 396
Cuidado residencial de bajo nivel (CRBN), 396
Curva normal, 128-130
áreas, 126
determinación, 128-130
escala tipificada, 126-127
con media y DT, 125
función densidad, 125
lectura inversa, 130-132
Curvas de densidad, 99
interpretación, 99
paradoja del continuo, 100
probabilidades, 100
- D**
- Datos categóricos
hipótesis compuestas, 355
intervalo de confianza:
de la proporción poblacional, 342-343
estudio de planificación, 346-347
unilateral, 345
niveles de confianza, 348
procedimiento de ajuste
chi-cuadrado, 349-352
distribución chi-cuadrado, 353
estadístico chi-cuadrado, 352-353
proporción muestral ajustada por Wilson, 337-340
dependencia del tamaño de la muestra, 340-341
distribución en el muestreo, 337
error típico (ET), 343
relación con la inferencia estadística, 340
resumen de métodos de inferencia, 360
resúmenes univariados, 51
variables dicotómicas:
alternativa direccional, 356-357
conclusión direccional, 356
Datos pareados, 400
Datos no normales, transformaciones para, 559
Datos perdidos, 21
Desviación típica (DT), 60, 175
estimación a partir del histograma, 63
interpretación, 60
regla empírica, 63
representación, 62
Desviación típica conjunta, 425
CM(intra), 426
CM(entre), 427
gl(intra), 426
SC(intra), 426
Diagrama de barras, 28
apiladas, 52
impresión visual de distribuciones, 389
Diagrama de caja modificado, 49
Diagrama de dispersión, 54
Diagrama de Venn, 94
Diagramas de barras apilados, 52
Diagramas de caja, 44, 46, 53
cuartiles, 44-46
IQR, 45-46
modificados, 49-51
uno al lado del otro, 53
Diagramas de puntos, 30
Diferencia de significación, 261, 494
Diferencia Significativa Honesta (DSH), 474
Diferencia Significativa Honesta (DSH) de Tukey, 474
Diferencia Significativa Mínima (DSM), 468
Dígitos significativos, 578
Diseño de bloques aleatorizados, un factor, 441
agrupamiento por bloques dentro de los sujetos, 442
cuadrados medios entre bloques, 448
gl(bloques), 448
SC(bloques), 448
test *F* de bloques completos aleatorizados, 446
Diseño pareado, 300, 311
aleatorizado, diseño completamente aleatorizado frente a, 314-315
análisis de datos, 315
ejemplos de, 311-313
experimentos con parejas de unidades, 311
limitaciones, 327-330
medidas repetidas, 312
objetivos del emparejamiento, 313-314
Diseño por bloques, 442. *Véase también*
Diseño por bloques aleatorizados, un factor
estudio de campo en agricultura, 442, 443
procedimiento de aleatorización, 444
Distribución binomial, 106, 339, 573-574
ajuste a datos, 114-117
aplicación al muestreo, 113
aproximación por la normal, 164, 165
coeficiente binomial, 109
fórmula, 109
ilustración, 107-109
media y desviación típica, 112, 575
modelo de pruebas independientes, 107
Distribución chi-cuadrado (χ^2), 353
Distribución de la aleatorización, 250
Distribución del rango estudiantizado, 474
Distribución en el muestreo, 147, 149.
Véase también Variabilidad del muestreo
media muestral, 149, 154
proporción muestral, 337-338
relación con la inferencia estadística, 150-151
y análisis de datos, 214
Distribución normal, 122
error de medida, 124
Distribución nula, 276-277
de la distribución chi-cuadrado, 353

- del test de signos, 320
 - estadístico de prueba, 317
 - Wilcoxon-Mann-Whitney, 286
 - Distribución *t* de Student, 180-181
 - condiciones, 272
 - consecuencias de su uso inapropiado, 273
 - resumen de la mecánica del test *t*, 275-276
 - verificación de condiciones, 273
 - Distribución χ^2 , 353
 - Distribuciones condicionales, 510-511
 - Distribuciones de frecuencias, 28
 - agrupadas, 31, 32
 - colas de, 32
 - efecto de transformaciones lineales, 67-68
 - mortalidad infantil, 29
 - Distribuciones *F*, 433
 - formas, 34-36
 - parámetros, 433
 - Distribuciones poblacionales, 197
 - condicionales, 510
 - condiciones, 198
 - de diferencias, 307
 - desviación típica, 156
 - media, 156
 - variable distribuida, 153
 - DSM de Fisher, 468
 - cálculos intermedios, 470
 - fórmula para cálculo, 470
 - tasa de error Tipo I por experimento, 471
 - DT. Véase Desviación típica (DT)
- E**
- Efecto principal, 454*n*
 - Error de medida, 124
 - Error de muestra debido al azar, 19
 - Error de muestreo, 19, 147. Véase también Magnitud del error no debido al muestreo
 - Error de Tipo I, 241, 420, 477
 - análisis de consecuencias, 242-243
 - riesgo, 280
 - Error de Tipo II, 241
 - análisis de consecuencias, 242-243
 - probabilidad, 242
 - riesgo, 280
 - Error no debido al muestreo, 20. Véase también Error de muestreo
 - muestreo aleatorio estratificado, 17
 - Error típico (ET), 174-175
 - combinación lineal, 461-462
 - grupos de personas, 178
 - desviación típica (DT) frente a, 175
 - proporción muestral ajustada mediante Wilson, 343
 - Error típico de la media, 175
 - Escala de densidades, 99
 - Escala estandarizada, 126
 - Espacio muestral, 94
 - Estadística descriptiva, 39
 - media, 40
 - mediana, 39
 - robustez, 41
 - Estadístico(s), 1, 26, 39, 74, 76
 - computador, 6
 - chi-cuadrado, 352-353, 369-370
 - de contraste, 227
 - descriptivos, 39-40
 - estadístico *t*, 227-228
 - Estimación de la media poblacional, 190
 - error típico (ET), 191
 - Estimación estadística, 173
 - desviación típica, 173, 174
 - media, 173, 174
 - notación, 174
 - Estratos, 18
 - Estrógeno equino conjugado (EEC), 186
 - Estructura de muestreo, 16
 - Estructura jerárquica, 193
 - Estudios ciegos, 10
 - Estudios observacionales, 7, 244, 245, 312
 - asociación espuria, 249
 - confusión, 246-249
 - dificultades, 246
 - frente a estudios experimentales, 245
 - ET. Véase Error típico (ET)
 - Evaluación de la normalidad, 133
 - gráfica de probabilidad normal, 134
 - funcionalidad, 135-136
 - toma de decisiones, 136-139
 - transformación para datos no normales, 139-140
 - Experimento, 9, 244, 245
 - doble ciego, 11
 - Extrapolación, 514
- F**
- Factor de corrección de población finita, 153
 - Factores, 452
 - aditivos, 454
 - Flujo de sangre miocárdial, 300
 - Formas de distribuciones, 34-36
 - bimodalidad, 34
 - unimodal, 34
 - Fórmula de mínimos cuadrados, 583-584
 - Frecuencia, 28
 - esperada, 352
 - en tablas de contingencia, 370
 - en un test chi-cuadrado, 389
 - observada, 352
 - relativa, 31
 - acumulada, 87
 - apilada, 52
 - histograma, 98
 - Fuentes de variación, 36
 - ALT en el suero, 37
 - Función de respuesta logística, 544
 - Función densidad, 125
- G**
- gl(intra), 426
 - gl(entre), 427
 - gl(total), 429
 - gl. Véase Grados de libertad (gl)
 - Grados de libertad (gl), 61, 181, 183
 - denominador, 433
 - dentro de los grupos (gl(intra)), 426
 - entre grupos (gl(entre)), 427
 - numerador, 433
 - Grados totales de libertad (gl(total)), 429
 - Gráfica de interacción, 454
 - Gráfica de probabilidad normal, 134-136
 - Gráfica de promedios, 502
 - Grupos de control, 11-12
- H**
- Hipótesis
 - alternativa, 226, 277
 - direccional, 252, 385
 - error de Tipo I, 241-243
 - error de Tipo II, 241-243
 - nula, 226
 - prueba de, 226
 - situaciones de aparición de errores, 241
 - prueba estadística, 227
 - test de aleatorización, 224
 - test *t*, 226, 227
 - Hipótesis alternativa direccional, 252, 356-357, 371

- en test de rangos con signo de Wilcoxon, 324
 - en test de signos, 319
 - en test de Wilcoxon-Mann-Whitney, 283-284
 - frente a alternativas no direccionales, 254-256
 - reglas, 257
 - test chi-cuadrado de bondad de ajuste, 356
 - Hipótesis de investigación. *Véase* Hipótesis alternativa
 - Hipótesis nula, 226, 372. *Véase también* Hipótesis alternativa
 - compuesta, 355
 - global, 422-423
 - Histograma, 29. *Véase también* Diagrama de barras
 - distribución de CFK, 34
 - estimación de la DT, 63-64
 - frecuencia relativa, 98
 - interpretación de las áreas, 34
- ## I
- Inferencia, 547
 - condiciones, 524
 - correlación, 491
 - de proporciones, 349
 - estadística, 71
 - Inferencia estadística, 71, 173
 - población, 73
 - respecto a β_1 , 516
 - error típico, 516
 - hipótesis nula, formulación, 518-520
 - implicación en el diseño, 517
 - intervalo de confianza, 517
 - Interacción, 454, 465
 - Interpolación, 513
 - Interpretación como densidad, 99
 - Interpretación de la definición de s , 60-61
 - Interpretación de la probabilidad como frecuencia, 86-88
 - Interpretación de la significación estadística:
 - diferencia significativa frente a diferencia importante, 261-262
 - intervalos de confianza, 263-265
 - tamaño del efecto, 262-263
 - Interpretación paramétrica de la regresión
 - distribuciones condicionales, 510
 - modelo de submuestreo aleatorio, 513
 - modelo lineal, 511
 - interpolación en, 513
 - predicción y, 514
 - poblaciones condicionales, 510
 - Intersección, 94
 - Intervalo de confianza, 184, 303-304, 462-463, 582-583
 - de diferencia de medias poblacionales, 207
 - cálculo de los grados de libertad, 207
 - condiciones de validez, 211
 - construcción del intervalo de confianza, 207-211
 - de medias poblacionales, 179
 - analogía del hombre invisible, 179-180
 - condición del método de la t -Student, 198
 - condiciones de validez, 196-198
 - determinación del valor crítico, 181
 - distribuciones t -Student, 180-181
 - verificación de las condiciones, 198
 - de proporciones poblacionales, 342-347
 - error típico de $\tilde{P}$, 343
 - intervalo de confianza del 95 % de p , 343-344
 - otros niveles de confianza, 345-346
 - planificación de un estudio para estimar p , 346-347
 - de puntuación, 582
 - de Wald, 582
 - de Wilson, 344, 583
 - relación, 187
 - unilateral, 187, 345
 - y aleatoriedad, 184
 - Intervalo intercuartílico (IQR), 45, 58, 62, 64
 - IQR. *Véase* Intervalo intercuartílico (IQR)
- ## J
- Jewett, Geoff, 180
- ## L
- Lactato deshidrogenasa, 261
- ## M
- m-clorofenilpiperacina (mCFP), 304
 - Media, 40, 103
 - desviaciones, 41
 - versus* mediana, 43
 - Media «ajustada en edad», 460
 - Media aritmética. *Véase* Media
 - Media global, 424
 - Media muestral, 41, 151
 - dependencia del tamaño de la muestra, 155-156
 - desviación típica, 152, 153, 157
 - distribución al muestreo, 151-152, 153, 157
 - forma, 153
 - Teorema Central del Límite, 155
 - Mediana, 39-40
 - distribución, 44-45
 - media frente a, 43
 - muestra, 75
 - visualización, 42
 - Medidas de dispersión, 58-66
 - coeficiente de variación, 61
 - comparación de, 64, 66-65
 - desviación típica (DT), 58-59
 - estimación a partir de un histograma, 63-64
 - desviación típica, 58
 - interpretación de la definición de s , 60-61
 - rango, 58
 - representación, 62-63
 - Metaestudio, 148
 - de un test t , 240
 - visualización de la distribución en el muestreo, 150
 - Método de Bonferroni, 473-474
 - ajuste de Bonferroni, 473
 - ventaja, 475
 - Método de Satterthwaite, 207*n*
 - Método de Welch, 207*n*
 - Métodos de inferencia, 557
 - diagrama de puntos, 559
 - Métodos no paramétricos, 559
 - Mínimos cuadrados. *Véase* Recta de regresión ajustada
 - Moda, 32
 - Modelo ANOVA de bloques
 - aleatorizados, 444
 - visualización de los efectos de los bloques, 445
 - Modelo de pruebas independientes, 107
 - Modelo de submuestreo aleatorio, 513, 524
 - Modelo lineal, 511, 536

estimación, 513
 interpolación en, 513
 predicción y, 514
 Monoamina oxidasa (MAO), 4, 177, 435

Muestra(s), 14
 aleatoria, 18, 147
 por estratos, 18
 procedimiento de selección, 16
 simple, 15
 por grupos, 17
 sesgada, 15

Muestreo aleatorio, 14
 aspectos prácticos, 17
 bivariado, 490-491, 524
 error de muestreo, 19
 métodos no simples, 17
 muestra, 14, 15
 aleatoria simple, 15
 sesgada, 15
 población, 14
 por estratos, 17
 por grupos, 17
 selección de la muestra aleatoria, 16
 sesgo de muestreo, 19
 simple, 15
 uso de la aleatoriedad, 16

N

Nivel de significación, 230
 Niveles, factor, 456
 Normal tipificada, 126

O

Operación aleatoria, 85, 106, 149
 lanzamiento de dado, 102
 lanzamiento de moneda, 85, 86
 Outliers, 47, 522
 crecimiento de rábanos en luz, 48-49
 valla inferior, 48
 valla superior, 48
 Oxigenación mediante membrana extracorpórea (OMEC), 384

P

P valor, 229
 Parámetro, 39*n*, 74, 76
 Perspectiva
 bayesiana, 280
 frecuentista, 280
 Placebo, 9
 Población, 14, 73
 correlación, 490

descripción, 74
 DT, 75
 hojas de tabaco, 75
 media, 75
 parámetro, 74
 variable categórica, 74-75
 Poblaciones condicionales, 510
 Potencia, 243, 267
 cálculo, 579-580
 dependencia con $(\mu_1 - \mu_2)$, 268
 dependencia con α , 267
 dependencia con n , 267
 dependencia con s , 267
 distribuciones normales, 269
 estudio de planificación, 268
 Precisión en la predicción, 532
 cálculo de intervalos, 534
 intervalos de confianza y de predicción, 532-533
 Predicción, 547
 Probabilidad, 84
 combinación, 90
 condicional, 368-369
 de error Tipo I por experimento, 468
 interpretación como frecuencia, 86
 operación aleatoria, 85
 experimento de lanzamiento de moneda, 85
 reglas, 94
 básicas, 94-95
 multiplicación, 96-97
 suma, 95-96
 Probabilidades de error
 análisis de pruebas médicas, 279
 árbol de probabilidad, 279
 interpretación, 278
 resultados hipotéticos, 280
 Proporción muestral ajustada por Wilson, 337-338
 dependencia del tamaño de la muestra, 340-341
 distribución en el muestreo, 338
 error típico (ET) de, 343
 intervalo de confianza, 342-344
 unilateral, 345
 niveles de confianza, 348
 planificación del estudio:
 para estimar p , 346
 sin información, 346
 relación con la inferencia estadística, 340
 Proteína de choque térmico (PCT), 564
 Prueba de Shapiro-Wilks, 140-141

Prueba sin distribución, 281
 Punto de influencia, 522
 efecto en el coeficiente de correlación, 523

R

Rango, 58
 Razón de ventajas, 403, 404
 diseño de control de caso, 406
 error típico (ET), 408, 409
 intervalo de confianza, 408-409
 Recta de DT, 498-500
 Recta de regresión ajustada, 487, 497, 546
 coeficiente de determinación, 506-507
 criterio de mínimos cuadrados, 504
 desviación típica residual, 505
 ecuación, 501
 fórmulas de mínimos cuadrados, 583
 gráfica de promedios, 502
 recta de regresión de mínimos cuadrados, 501
 suma de cuadrados residual, 503
 Recta de regresión, 56
 de mínimos cuadrados, 501
 Regla empírica, 63-64
 Reglas de la suma, 95-96
 Reglas de multiplicación, 96-97
 Regresión curvilínea, 539
 Regresión lineal y análisis de correlación, 485-554
 análisis de covarianza, 540-541
 coeficiente de correlación, 487-494
 definición, 487
 fórmulas, 546
 inferencia estadística con respecto a correlación, 516-520
 intervalo de confianza de ρ , 492.493
significativo, uso del término, 494
 coeficiente de determinación, 506-507
 ejemplos de, 487, 490-491, 493
 fórmulas de inferencia, 547
 inferencia estadística con respecto a β_1 , 516-520
 contraste de las hipótesis, 518-520
 error típico de β_1 , 516-517
 intervalo de confianza de β_1 , 517
 modelo de muestreo aleatorio bivariado, 490

- modelo lineal, 510-515
 - constancia de la desviación típica, 511
 - definición, 511-512
 - distribuciones condicionales, 510
 - estimación de, 513
 - gráfica de promedios, 502-503
 - interpolación en, 513-514
 - linealidad, 511
 - modelo de submuestreo aleatorio, 513
 - poblaciones condicionales, 510
 - y predicción, 514
 - pautas de interpretación, 521-529
 - condiciones de diseño, 517, 524
 - condiciones de distribución poblacional, 524
 - condiciones de muestreo, 524
 - condiciones de parámetros, 524
 - condiciones para inferencia, 524
 - correlación y causalidad, 493
 - descripciones inadecuadas de conjuntos de datos, 521-524
 - gráficas de residuo, 527-528
 - modelo lineal y condición de normalidad, 526
 - transformaciones, uso de, 528-529
 - recta de regresión ajustada, 497-510
 - criterio de mínimos cuadrados, 504
 - desviación típica residual, 505-506
 - ecuación de la recta de regresión, 501
 - fórmulas, 546
 - recta de mínimos cuadrados, 504
 - recta de regresión, 501-502
 - suma de cuadrados del residuo, 503-504
 - regresión logística, 541-546
 - regresión y correlación no paramétricas y robustas, 540
 - regresión y test t , 535-539
 - resumen de fórmulas, 546-547
 - Regresión logística, 541, 542
 - Regresión y correlación
 - análisis de covarianza, 540
 - causas de descripción inadecuada, 521-522
 - condiciones de inferencia, 524
 - etiquetado de X e Y , 526
 - extensiones de mínimos cuadrados, 539
 - gráficas de residuos, 527
 - interpretación, 521
 - modelo lineal y condición de normalidad, 526
 - múltiple, 539
 - no paramétrica y robusta, 540
 - pautas de condiciones de muestreo, 524
 - regresión logística, 541
 - relación curvilínea con X , 521
 - test t , 535
 - uso de transformaciones, 528
 - Residuo, 437, 446
 - desviación típica, 505
 - gráficas, 527, 528, 531
 - Respiración del suelo, 281-282
 - Resumen de cinco números, 46
 - Resumen de univariado, 51
 - Riesgo relativo, 402-403
 - Robustez, 41-42
- S**
- SC(intra), 426
 - SC(entre), 427
 - SC(residuo), 504
 - SC(total), 428
 - Sesgada hacia la derecha, 32
 - Sesgo, 19
 - de muestreo, 19
 - de panel, 12
 - efecto, 42
 - moderado, 273
 - selección, 73
 - sin respuesta, 21
 - Suma de cuadrados de los residuos (SC(resid)), 504
 - Suma de cuadrados dentro de los grupos (SC(intra)), 426
 - Suma de cuadrados entre grupos (SC(entre)), 427
 - Suma de cuadrados total (SC(total)), 428
 - Sustitución doble, 149
- T**
- Tabla bivariada de frecuencias, 51
 - Tablas de contingencia, 367
 - Tabla de contingencia $r \times q$, 389
 - contextos, 391-392
 - test chi-cuadrado, 389
 - condiciones de validez, 394
 - consideraciones sobre la potencia, 396
 - frecuencias esperadas, 390
 - verificación de las condiciones de diseño, 394
 - Tablas de contingencia 2×2 , 367.
 - Véase también Tabla de contingencia
 - contextos, 376
 - datos pareados, 400
 - test de McNemar, 401
 - transmisión de VIH, 400-401
 - descripción verbal de la asociación, 380
 - estadístico chi-cuadrado, 369
 - frecuencias esperadas, 370
 - frecuencias observadas, 369
 - hechos sobre filas y columnas, 379-380
 - independencia y asociación, 376
 - notas de cómputo, 372
 - procedimiento de contraste, 371-372
 - razón de ventajas, 403, 404-408
 - riesgo relativo, 402-403
 - test chi-cuadrado, 368
 - hipótesis nula, 368
 - intervalo de confianza, 397-399
 - probabilidad condicional, 368
 - relación con el test, 399
 - Tamaño del efecto, 262-263
 - Tasa de error de Tipo I, 468
 - por experimento, 468, 478
 - Teorema Central del Límite, 153, 155, 161, 344
 - y aproximación normal a la distribución binomial, 577
 - Terapia médica convencional (TMC), 384
 - Test chi-cuadrado, 352, 368
 - tabla de contingencia 2×2 , 368, 371
 - tabla de contingencia $r \times k$, 389
 - test exacto de Fisher, 384
 - Test chi-cuadrado de bondad de ajuste, 352, 354
 - estadístico chi-cuadrado, 352
 - hipótesis nula compuesta, 355
 - diagramas de barras, 351
 - variables dicotómicas, 356
 - alternativa direccional, 356-357
 - conclusión direccional, 356
 - Test de aleatorización, 221-226, 287
 - Test de asociación. Véase Test chi-cuadrado
 - Test de independencia. Véase Test chi-cuadrado

- Test de Mann-Whitney. Véase Test de Wilcoxon-Mann-Whitney
- Test de McNemar, 401
distribución chi-cuadrado, 401
transmisión de VIH para análisis de niños, 401
- Test de rangos ordenados de Wilcoxon, 322. Véase también Test de Wilcoxon-Mann-Whitney
acotación del P valor, 324
alternativa direccional, 324
 cálculo del valor absoluto, 323
 rangos con signo, 324
 valores críticos, 324
aplicabilidad, 324-325
tratamiento de los ceros, 324
tratamiento de los empates, 324
- Test de signos, 316, 325. Véase también Test t
alternativa direccional, 319
aplicabilidad, 320-321
cálculo del valor crítico, 320
distribución nula, 319-320
 obtención del P valor, 317
 tiempos de supervivencia, 317
 valores críticos, 318
horquilla del P valor, 319
tratamiento de los ceros, 319
- Test de Wilcoxon-Mann-Whitney, 273, 280, 580-581. Véase también Test de rangos con signo de Wilcoxon
alternativa direccional, 284
aplicabilidad, 281-281
condiciones, 286
direccionalidad, 283
distribuciones nulas, 285
frente a test de aleatorización, 286
frente a test t , 286-287
métodos, 282
 P valores, 285
planteamiento de H_0 y H_A , 281
razones, 284
- Test exacto de Fisher, 384
comparación con el test chi-cuadrado, 386
hipótesis alternativa, 385
y alternativas no direccionales, 386-388
- Test F , global, 433
distribuciones F , 433
estadístico F , 433
y test t , 434
- Test no paramétrico, 281
- Test t , 224, 226, 463. Véase también Test de signos
condiciones, 272-273
de dos muestras, 559
estadístico t , 227-228
hipótesis alternativa, 226, 277
hipótesis nula, 226, 277
metaestudio de, 239
 P valor, 229, 277-278
 conservativo, 232
 de dos colas, 229, 251
 determinación, 232
 extracción de conclusiones, 230-232
 interpretación, 239-241, 278-280
 nivel de significación frente a, 241
potencia, 243
presentación de resultados, 233
relación entre el test y el intervalo de confianza, 237-239
- Test t de una cola, 252-254
hipótesis alternativas direccionales, 252
 frente alternativas no direccionales, 254-256
 regla, 257
 P valor, 253
procedimiento de test, 252-253
- Tomografía por emisión de positrones (TEP), 300n
- Toque terapéutico (TT), 565
- Transformación aditiva, 68-69
- Transformación de Fisher, 492
- Transformación multiplicativa, 67
- Transformaciones, 528
efecto de, 66-70
lineales, 66-69
 codificación, 67
 distribución de frecuencias, 67-68
 efecto sobre la distribución de frecuencias, 67-68
multiplicativa, 67
no lineales, 69-70
- U**
- Unidades observacionales, 27
 anidadas, 193
 notación, 27
- Unimodal, 34
- V**
- Valor ajustado, 438
- Vallas, 48
- Variabilidad del muestreo, 147, 149.
 Véase también Aspectos de probabilidad
 metaestudio, 148
- Variable aleatoria, 102
binomial, 106, 107-108
 desviación típica (DT), 112
 fórmula de la distribución, 99
 media, 112
continua, 102
discreta, 102
 media, 103
 varianza, 103
reglas, 105
- Variable(s), 26
aleatoria, 102
categórica, 26
continua, 26
de respuesta, 244, 441n
dicotómica, 356
discreta, 27
efecto de las transformaciones, 66
 transformación aditiva, 68-69
 transformación multiplicativa, 67
 transformaciones lineales, 66
explicatoria, 244
indicadora, 536
notación, 27
numéricas, 26
ordinales, 26
relaciones:
 relaciones categórica-categórica, 51
 relaciones numérica-categórica, 53
 relaciones numérica-numérica, 54
- Varianza
agrupada, 204
análisis de un factor, 423-424
análisis del modelo, 431-432
conjunta, 425, 536
variable aleatoria, 103

ÍNDICE DE EJEMPLOS

- A**beraciones cromosómicas, 465
Abultamientos cromosómicos, 564-565
Ácido giberélico, 559-561
Alanina aminotransferasa (ALT) en el suero, 37
Alas de mariposa, 173-174, 175, 182-183, 191
Albinismo, 107-108
Alcohol y MOPEG, 73
Alfalfa y lluvia ácida, 441, 445, 448-449
ALT en el suero, 37
Altura(s)
 de 11 mujeres, 135-136
 de estudiantes, 33
 de hombres, 102
 de personas, 268, 269
 de plantas y resistencia a enfermedades, 378-379, 380
 y pesos de niñas, 61-62
Anatomía canina, 193
Anfetamina y consumo de alimentos, 485-486, 502, 507, 510-511, 514
Anidamiento de chorlitos, 389, 390-391
Ardillas, 307
Área de hojas, 224-225
Arsénico en arroz, 486-487, 498-500, 501, 504, 505, 506, 507, 513-514, 533
Asma bronquial, 10
Ataques al corazón y aspirina, 409
Autismo, 9
Ayudas sobre el aborto, 21
Azúcar e hiperactividad, 22
- B**acterias y cáncer, 2-3
Biorrealimentación y presión sanguínea, 327-328
Bloques
 en un estudio agrícola de campo, 442
 por camadas, 441-442
- C**aballitos del diablo, 569
Calcio en las plaquetas y presión sanguínea, 491-492, 493, 518-519, 538-539
Cáncer
 de esófago, 542-546
 de pecho, 344-345
 de pulmón, 75
 y fumar, 312, 403, 404, 405-409
 y bacterias, 2-3
 y tinte para el pelo, 252
Cangrejos de arena, 18-19
Capacidad vital, 202-203, 204, 205
 forzada (FVC), 459-460, 462
Cardo La Graciosa 17-18
Células nerviosas
 densidad de, 322-324
 tamaños de, 19
 tiempos entre disparos de, 123-124
Cerdas, tamaño de camada de, 29-30
CFK en el suero, 31-32
Clofibrate, 11
Colesterol en el suero, 122, 133-134, 152-153
 medida, 328-329
 y glucosa en el suero, 524-525
 y presión sanguínea, 539-540
Colitis ulcerosa, tratamiento de la, 20
Color
 de cabello y de ojos, 95, 96, 97, 376-377, 380-381, 391-392
 de flores de pascua, 28, 31
Condición de zurdo, 346, 347
Consumo de cigarrillos, 245
Contaminación de *E. coli* entre cuencas, 51, 52
Contenido de humedad, 134
Cosecha de tomates, 264, 265
Creatina fosfoquinasa (CFK) en el suero, 31-32
Crecimiento
 bacteriano, 149
 de crisantemos, 59, 60, 61
 de lentejas, 139, 141
 de novillos, 525
 de plantas de soja, 452-453, 456-457, 460-461, 462-463, 464, 528-529
 de rábanos, 47, 53-54
 con luz, 48
 de virus, 312-313, 318-319
- D**ado, 102, 103, 104
Densidad mineral del hueso, 186
Diámetros de árboles, 101
Dolencia de la arteria coronaria, 12-13
Dolor
 de cabeza, 249
 de cuello y mochilas escolares, 28
- E**jercicio y triglicéridos en suero, 312
Embarazo, fumar durante, 343
Error de medida, 124
Espesor de cáscara de huevo, 123, 184
Esquizofrenia y MAO, 4, 177-178
Estrellas de mar, 567
Estrógeno y esteroides, 568
Estudio de campo agrícola, 443, 444
- F**acetas de ojos, 161
Falsos positivos, 93
Fertilizante para berenjenas, 311, 315
Festival de la Luna de la Cosecha, 357
Fisioterapia, 396-397
Flexibilidad, 221-222
Flores de pascua, color de, 28, 31
Flujo sanguíneo, 300, 302-304
Fuego y hábitat de ciervos, 349-351, 351-352, 353, 354, 355
Fumar
 durante el embarazo, 343
 y cáncer de pulmón, 312, 403, 404, 405-409
 y peso al nacer, 248
- G**anancia
 de peso de corderos, 39-40, 424, 425, 426, 427, 429, 432, 434, 438
 diaria de ganado, 62
Gatos mutantes, 107, 110
Gemelos, 567-568

Germinación de esporas, 193-194
 Glucosa en sangre, 99, 100
 Grillos, tiempos de canto de, 42, 69-70
 Grupo sanguíneo, 71-72, 73, 94, 97, 111-113

Hábitat de ciervos y fuego, 349-352, 353, 354, 355
 Hematocrito en hombres y mujeres, 244
 Hiperactividad, azúcar e, 22
 Hojas de tabaco, 75-76

Inflamación de tejidos, 273-275
 Injertos de piel, 316-317
 Inmunoterapia, 242
 Inundación y ATP (Adenosina trifosfato), 3

Lanzamiento de moneda, 85, 86, 89-90
 Larvas de insecto, selección de alimento por, 5-6, 94-395
 LD en el suero, 261-263, 264
 Ligadura de arteria mamaria, 10
 Longitud
 de cangrejos de río, 238-239
 de peces, 19, 128-129, 131
 y peso de serpientes, 487, 490, 513, 516, 518
 Luz y sordera, 7

Maíz dulce, 419-420, 438, 439
 MAO y esquizofrenia, 4, 177-178
 Marihuana
 e inteligencia, 192-193
 y glándula pituitaria, 242
 Masa, 105
 Matrimonio y salud, 8
 Medicaciones, 102
 Metabolismo del nitrato, 20
 Microfósiles, 35
 Migrañas, 366-367, 368, 369-370, 371-372, 398-399
 Mochilas escolares y dolores de cuello, 28
 Monoamina oxidasa (MAO) y esquizofrenia, 4, 177-178
 MOPEG y alcohol, 73
 Mortalidad infantil, 29
 Moscas de la fruta, muestreo de, 85, 86-87, 90, 112
 Muestras del suelo, 568
 Muestreo de moscas de la fruta, 85, 86-87, 90, 112
 Música y caléndulas, 239-240, 257

OMEC, 345, 384, 385-386
 Orientación sexual, 7
 Ostras y pastos marinos, 468-470, 471-472, 473-474
 Óxido nítrico, 90-92

Pargilina y consumo de sacarosa, 244-245
 Pastos marinos y ostras, 468-470, 471-472, 473-474
 Peces, longitudes de, 19, 128-129, 131
 Pérdida de peso postparto, 270
 Perros, toxicidad en, 8-9
 Peso, 67
 corporal, 262, 263, 264
 de corderos al nacer, 175-176, 177
 de semillas, 37, 154, 155-156, 157
 del cerebro, 37
 del tórax de mariposas, 210-211
 y alturas
 de hombres jóvenes, 511, 512, 526
 de niñas, 61-62
 Pies a pulgadas, 105
 Plantas
 de avena, 74
 rápidas, 208, 230-231, 232-233
 Polinización de flores, 395-396
 Presión sanguínea, 45, 58
 en ratas, 149
 y biorrealimentación, 327-328
 y calcio en las plaquetas, 491-492, 493, 518-519, 538-539
 y colesterol en el suero, 539-540
 Prevención del uso de tabaco, 569
 Prueba médica, 92, 279
 Pulsaciones después de hacer ejercicio, 64
 Pulso, 45-46, 48

Quimioterapia y THC, 321

Raza y tamaño del cerebro, 346-247
 Rendimiento de sedimentos, 199
 Reproducción de un alga, 493
 Resfriado común, 11-12
 Resistencia a los hongos del maíz, 20
 Respiración de la tierra, 281, 282-283

Sacarosa en raíces de remolacha, 19
 Salud y matrimonio, 8
 Sapos, 457-458
 Selección de alimento por larvas de insecto, 5-6, 394-395

Selenio en ballenas, 54-55
 Semillas
 de lino, 354-355
 por fruto, 186-188
 Serpientes, longitud y peso de, 487, 490, 513, 516, 518
 Sexos de niños, 114-117
 Soda contaminada, 337, 338, 339-340
 Sordera y luz, 7
 Suplemento
 de niacina, 252, 253-254, 255
 de hierro en bebidas de fruta con leche, 455, 456
 Sustitución de rodilla, 149-150

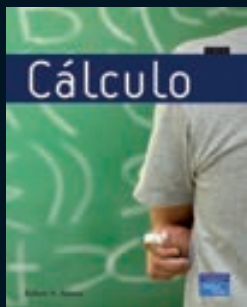
Tamaño
 corporal y gasto de energía, 6
 de cabeza de orugas, 540
 de camada de cerdas, 29-30
 de familia, 102
 de la mano, 97
 Tamoxifeno, 563-564
 Temperatura, 105
 corporal, 67-68
 del océano, 497-498
 Test de VIH, 21, 367, 373
 THC y quimioterapia, 321
 Tiempo
 de reacción, 162-164
 entre disparos de células nerviosas, 123-124
 entre descargas de células, 35-36
 Tinte para el pelo y cáncer, 252
 Tolueno y cerebro, 226, 228, 229, 230, 535-538
 Tonsilectomía, 204
 Toque terapéutico, 565-567
 Toxicidad en perros, 8-9
 Transmisión del VIH a niños, 400-401
 Tratamiento del acné, 329
 Triglicéridos en el suero, ejercicio y, 312

Ultrasonidos, 249

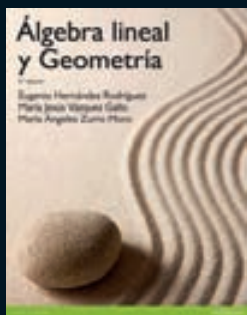
Vacuna
 de la gripe, 387-388
 para el ántrax, 2
 Vacunaciones, 568
 Valoración del hambre, 304-307
 Varicela, 113
 Velocidad de nado de ballenas, 561-563

Este libro es un texto introductorio en Estadística, dirigido específicamente a estudiantes que se especializan en Ciencias. Sus principales objetivos son mostrar a los estudiantes cómo se utiliza el razonamiento estadístico en la investigación biológica, médica y agrícola; dotarles de la capacidad necesaria para realizar análisis estadísticos e interpretar los resultados; y destacar la importancia de aspectos estadísticos básicos como la aleatoriedad, las variables de confusión y la función de la replicación independiente.

Fundamentos de Estadística para las Ciencias de la Vida ha sido diseñado para cubrir la asignatura de Estadística que se ha implantado con los nuevos grados en primer curso para todas las carreras relacionadas con Salud y Medio Ambiente.



Cálculo 6.^a Ed.
Robert A. Adams
Concebido para cursos de Cálculo general, especialmente para los estudiantes de Ciencias e Ingeniería.
ISBN: 9788478290895



Álgebra lineal y Geometría 3.^a Ed.
Eugenio Hernández Rodríguez
María Jesús Vázquez Gallo
María Ángeles Zurro Moro
Esta obra cubre todos los temas de un primer curso de enseñanza universitaria en Álgebra lineal y Geometría, tanto en las facultades de Ciencias como en las escuelas de Ingeniería.
ISBN: 9788478291298

booksmedicos.org

PEARSON

www.pearson.es

ISBN 978-84-782-9141-0



9 788478 291410