

Capítulo 7:

Prueba de significación y contraste de hipótesis

Introducción

Este capítulo formaliza la respuesta a dos preguntas diferentes pero relacionadas: «¿Qué sé?» (inferencia) y «¿Qué hago?» (decisión). Se define la prueba y el valor de P en el entorno de la evidencia empírica o inferencia sobre conocimiento. Por su parte, los riesgos α y β y el contraste de hipótesis se enmarcan en la decisión entre dos acciones alternativas. Finalmente, se distingue entre planteamientos para poner a prueba diferencias y planteamientos de equivalencia.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Percibirá la mayor dificultad de interpretar una prueba de significación o un contraste de hipótesis que un intervalo de confianza.
- Sabrá que el intervalo de confianza excluye aquellos valores del parámetro que hubieran sido rechazados en una prueba de significación.
- Sabrá que la hipótesis precede a la recogida de datos.
- En una prueba de significación, escribirá en la hipótesis H la negación de su objetivo de investigación.
- Identificará valor de P con inferencia y lo interpretará como magnitud de evidencia en contra de la hipótesis H .
- En una prueba de significación, si no rechaza H , dirá: «Nada se opone a aceptar H » o «no se han encontrado diferencias» en lugar de «no existen diferencias».
- No interpretará el valor de P como el riesgo de haberse equivocado.
- Escribirá $P < 0,001$ en lugar de $P = 0,000$.
- Interpretará el contraste de hipótesis entre H_0 y H_1 como una decisión entre dos acciones alternativas (A_0 y A_1).
- En un contraste, distinguirá entre riesgo α y riesgo β .
- Identificará riesgos α y β con la frecuencia de decisiones erróneas.
- Calculará la prueba de una media para el caso de muestras grandes o de variables con distribución normal.
- Distinguirá entre un planteamiento para demostrar diferencias y otro para establecer equivalencia.
- Distinguirá entre equivalencia, no inferioridad y no superioridad.
- Interpretará el margen de equivalencia como diferencias no relevantes.
- Tendrá curiosidad por el planteamiento bayesiano.

Objetivos

A la vista de la información aportada por la muestra, las dos principales preguntas de la inferencia estadística son: ¿qué valores del parámetro son creíbles? Y ¿se puede negar cierto valor del parámetro? La primera, mediante intervalos de confianza, se resolvió en el capítulo anterior; la segunda se expone en éste.

Comentario



Dado que el parámetro representa un valor poblacional, la inferencia pretende alcanzar modelos universales o absolutos. Estos modelos deben ser interpretados de acuerdo con las condiciones y características del estudio (quizá las conclusiones sean válidas sólo para adultos voluntarios sanos), pero son universales en el sentido de que se aplican a toda su población. En resumen, ambos procedimientos (intervalos de confianza y prueba de significación) son ambiciosos, ya que pretenden formalizar el conocimiento científico.

La pregunta que responden los intervalos de confianza (¿qué valores son creíbles?) engloba, de alguna manera, a la pregunta de la prueba de significación (¿se puede negar cierto valor?). Los intervalos de confianza aportan más información y son más fáciles de entender, asimilar y explicar. ¿Qué interés ofrece, entonces, poner a prueba una hipótesis? Pues, simplemente, que puede ser la auténtica pregunta de interés.

Ejemplo 7.1



Saber si un fármaco es más eficaz que otro puede reducirse a conocer si la diferencia de sus medias en la respuesta de interés es o no es exactamente el valor 0. Por tanto, poder negar el valor 0 implica haber demostrado que un producto es más eficaz que otro.

Esta prueba sobre una hipótesis puede abordarse desde el punto de vista de la inferencia (¿qué sé?) o de la decisión (¿qué hago?), y se ven, respectivamente, en la prueba de significación, PS, y en el contraste de hipótesis, CH.

Inferencia: prueba de significación

Se desea **poner a prueba** o confrontar una hipótesis previa H con la información que proporcionan los datos.

Ejemplo 7.2



Desde hace algún tiempo, un residente se juega a cara o cruz las guardias que coinciden con las fiestas familiares. Su compañero lanza su moneda y... ¡siempre gana!, de for-

Ejemplo 7.2 (Cont.)

ma que la confianza en él se va deteriorando, hasta que un día se plantea estudiar formalmente si la moneda está apañada. Así, el problema consiste en analizar si es cierta la: Hipótesis $H : \pi = 0,5$ (moneda correcta), donde π representa la probabilidad de cara que se desea contrastar. Obsérvese que, para demostrar que la moneda está cargada, se pone en H que es correcta. Si se lanza $n = 100$ veces la moneda y se observa la proporción p de caras, se dispondrá de cierta «evidencia» empírica que conviene asimilar. Supóngase que se observa una proporción $p = 0,63 = 63\%$. Este resultado conduce a creer, de alguna manera, que la moneda está «cargada»: que no es cierto que $\pi = 0,5$. En cambio, si el resultado fuera $p = 0,52 = 52\%$, se consideraría «compatible» con que la moneda no esté cargada. Cuanto más se aleje p de $0,5$, hay más información en contra de la hipótesis H de que la moneda es correcta.

Comentario

La aleatoriedad asociada a este proceso hace que no esté libre de riesgos. Es posible que una moneda perfecta, no cargada, genere una observación de 63 caras en 100 lanzamientos. Y, de forma recíproca, también es posible que una moneda que no sea perfecta y que tenga una probabilidad de cara diferente de 0,5 genere una muestra con un 50% de caras.

Comentario

Por supuesto, se podría abordar el problema desde un punto de vista físico y, dando por bueno el conocimiento actual de esta ciencia, estudiar la composición de la moneda, su centro de gravedad, su circunferencia, etc. Ahora bien, sea cual sea la respuesta del estudio teórico, siempre conviene estudiar qué dicen los datos, no sea que convenga revisar el modelo teórico.

La hipótesis H establece una condición sobre el parámetro poblacional que se desea poner a prueba, confrontando ésta con la información que proporcionan los datos. Esta información se «condensa» en un estadístico apropiado, que fluctúa aleatoriamente según una distribución dependiente del verdadero valor del parámetro. Cuando H es correcta, la distribución es conocida, y el estadístico se localizará de forma previsible o, lo que es lo mismo, no se alejara mucho de una zona determinada. Por tanto, cuanto más lejos se encuentre el estadístico de la zona asociada a H , menos verosímil se presenta ésta, y más credibilidad cobra la posibilidad de que el estadístico proceda de otra distribución, es decir, de parámetros distintos al que determina H .

Definición

La prueba de significación es una técnica de inferencia estadística para juzgar si una propiedad que, se supone, cumple una población es compatible con lo observado en una muestra de la misma.

Ejercicio 7.1

En relación con la prueba de significación (elija una):

- Se desea conocer el valor de cierto parámetro.
- Se construye una hipótesis sugerida por los datos.
- Se busca «evidencia» (pruebas) a favor de la hipótesis H que se desea demostrar.
- Ninguna de las anteriores es correcta.

Ejercicio 7.2

Escriba la hipótesis para contrastar si un nuevo tratamiento es mejor que uno clásico:

- H : el nuevo tratamiento no es mejor que el clásico.
- H : el nuevo tratamiento es mejor que el clásico.
- H : el rendimiento del nuevo tratamiento supera al clásico.
- Ninguna de las anteriores es correcta.

Así pues, se requiere un proceso formal que permita «incorporar» la información muestral o «evidencia» empírica. Este proceso debe ser transparente, en el sentido de ser reproducible por un segundo investigador.

Valor de P

Un procedimiento consiste en calcular el valor de P (P value) o probabilidad de que se presente un valor del estadístico más alejado de la hipótesis H que el observado.

Recuerde

Si el valor de P (P value) es pequeño, se dice que H es inverosímil.

Recuerde

El valor de P puede interpretarse como «cuán inverosímil es el resultado observado si H fuera cierta» o «hasta qué punto el resultado observado es probabilísticamente compatible con H ». Lo que suele interpretarse, cuando P es pequeña, como que hay «suficiente evidencia o pruebas en contra de H » para creer que el resultado es estadísticamente significativo».

Mecánica de la prueba de significación

La prueba de significación se basa en el siguiente proceso formal (fig. 7-1):

- i) Se escoge una variable, que valora el objetivo del estudio.
- ii) Se elije un diseño y un estadístico que resuma la variable en la muestra.
- iii) Se especifica una hipótesis (H) que se desea poder rechazar.
- iv) Se especifica la distribución del estadístico bajo H y las premisas necesarias.
- v) Se realiza el experimento y se calculan los valores del estadístico y del valor de P.
- vi) Si el valor de P es muy pequeño se dice que H es inverosímil.
- vii) Se reporta el estimador con su IC_{95%}.

A continuación se explica este proceso para el caso de una probabilidad.

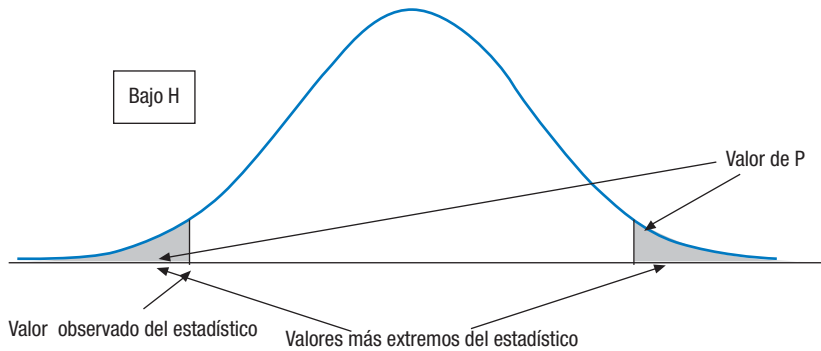


Figura 7-1 El valor de P indica la probabilidad de observar valores del estadístico igual o más extremos que el observado, en el caso de que H sea cierta.

Prueba de significación de una probabilidad

Se pueden utilizar los conocimientos sobre la distribución de la proporción p observada en una muestra para poner a prueba una hipótesis H sobre una probabilidad poblacional π .

Si la probabilidad π de la población origen de la muestra es 0,5, se escribe $H : \pi = 0,5$

Recuerde



Si la probabilidad poblacional de cierto evento es π , la distribución de la proporción observable p en muestras de tamaño n puede aproximarse por una distribución normal centrada alrededor de π con varianza $\pi(1-\pi)/n$:

$$p \rightarrow N(\pi, \pi(1-\pi)/n)$$

Comentario



Recuerde que las condiciones de aplicación de esta fórmula son que la muestra sea grande y la probabilidad π no extrema:

$$\pi \cdot n \geq 5 \quad \text{y} \quad (1 - \pi) \cdot n \geq 5$$

Comentario



Observe que, como π vendrá especificado por la hipótesis, ahora no es necesario sustituirlo por su estimación muestral p ni por el valor del producto máximo $\pi(1 - \pi)$.

En el ejemplo anterior de la moneda, con una muestra de $n = 100$,

i) Variable: resultado cara o cruz

ii) Estimador: proporción p de caras

iii) Hipótesis $H: \pi = 0,5$ (moneda correcta)

iv) Si H es cierta: $p \rightarrow N(\pi, \pi(1 - \pi) / n) = N(0,5, 0,05^2)$

Premisas: muestra grande $\pi \cdot n > 5$ y $(1 - \pi) \cdot n > 5$

v) Caso a) con $n = 100$ se observan 63 caras

$p = 63 / 100 = 0,63 = 63\%$

$\hat{Z} = (p - \pi) / \sqrt{\{\pi(1 - \pi) / n\}} = (0,63 - 0,5) / \sqrt{\{0,5 \cdot 0,5 / 100\}} \approx 0,13 / 0,05 = 2,6$

En la tabla 4-2 se obtiene que al valor 2,6 le corresponde una probabilidad unilateral de 0,0047, por lo que el valor bilateral de P es $0,0094 < 0,01$.

vi) Por tanto (fig. 7-2), como el valor de P (o probabilidad de observar un valor de p tan o más alejado de $H: \pi = 0,5$) es $P < 0,01$, se considera inverosímil que π es 0,5 con un valor de $P < 0,01$.

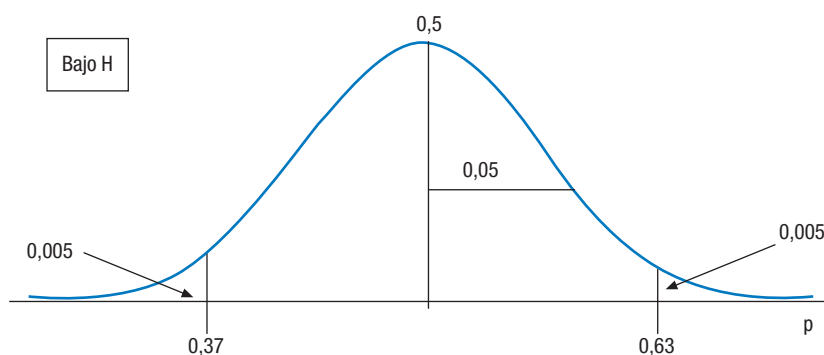


Figura 7-2 Bajo $H: \pi = 0,5$ y con una muestra $n = 100$, la distribución de p , proporción observada en la muestra, es Normal $(0,5, 0,05^2)$. Si se observan 63 caras, $p = 0,63$ y la probabilidad de observar 63 o más caras es de 0,005, que junto a la de observar 63 o más cruces (27 o menos caras) hace $P = 0,01$.

vii) $IC_{95\%}: p \pm Z_{\alpha/2} \sigma_p \approx 0,63 \pm 1,96 \cdot 0,05 \approx 0,63 \pm 0,10 = 0,53, 0,73$

La «auténtica» proporción de cara π se encuentra entre 53 y 73% (fig. 7-2).

v) Caso b) con $n = 100$ se observan 52 caras

$\pi = 52/100 = 0,52 = 52\%$

$\hat{Z} = (p - \pi) / \sqrt{\{\pi(1 - \pi) / n\}} = (0,52 - 0,5) / \sqrt{\{0,5 \cdot 0,5 / 100\}} \approx 0,02 / 0,05 = 0,4$

La tabla 4-2 proporciona, para el valor 0,4, una probabilidad unilateral de 0,3446, a la que corresponde $P = 0,6892$ bilateral.

vi) Por tanto (fig. 7-3), valor de $P = \text{Prob}(p(\text{cara}) \geq 0,52 \text{ y } p(\text{cara}) \leq 0,48) \approx 0,69$

Como el valor de P es grande, no es inverosímil, o más formalmente: «nada se opone a aceptar H ».

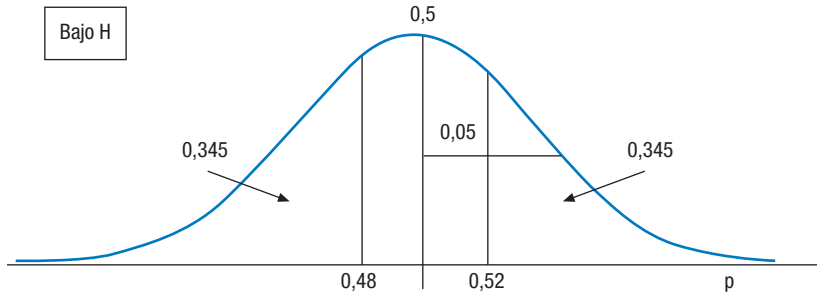


Figura 7-3 Si se observan 52 caras, $p = 0,52$ y la probabilidad de observar 52 o más caras es de 0,345, que junto a su simétrica (observar 52 o más cruces = 48 o menos caras) hace $P = 0,69$.

vii) $IC_{95\%}: p \pm Z_{\alpha/2} \sigma_p = 0,52 \pm 1,96 \cdot 0,05 \approx 0,52 \pm 0,10 = 0,42, 0,62$
La «auténtica» proporción de cara π se encuentra entre 42 y 62%.

Ejercicio 7.3



Una serie de 400 pacientes con sida han recibido, en diferentes etapas de su seguimiento, dos tratamientos diferentes, A y B que son, a priori, potencialmente similares. Ahora, se les ha preguntado cuál prefieren, y un 58% ha optado por el A. Se desea saber si es razonable creer que son igualmente preferidos. Escriba todos los pasos del proceso.

Ejercicio 7.4

Repita el ejercicio 7.3 asumiendo que las preferencias por A han sido 53%.

Un programa informático, al redondear, proporciona un nivel de significación con muchos ceros (0,000...) que parecería indicar un resultado imposible, de probabilidad nula. Como ello no es así, se expresa de la siguiente forma: $P < 0,001$.

Recuerde

Nunca escriba $P = 0,000$; en su lugar, ponga $P < 0,001$.

Ejercicio 7.5

Los usuarios de una biblioteca llevan años protestando por las prestaciones del sistema de búsqueda disponible para realizar sus consultas. Los responsables de la biblioteca deciden valorar la posibilidad de cambiar el sistema. Durante el período de prueba, han realizado un experimento comparando ambos sistemas mediante una escala que mide la satisfacción de los usuarios. Hacen la prueba anterior de comparación de preferencias y resumen sus resultados con la siguiente frase: *el nuevo sistema genera mayor satisfacción en los usuarios ($P < 0,01$)*. ¿Cuál o cuáles de las siguientes afirmaciones son ciertas?:

- Se rechaza la hipótesis H de que la satisfacción es igual en ambos grupos.
- Suponiendo que ambos sistemas generen la misma satisfacción, la probabilidad de haber obtenido un resultado tan o más extremo que el observado es menor del 1%.
- Creemos que el resultado observado refleja una diferencia poblacional, en el conjunto de todos los posibles casos, del nivel de preferencias.
- La proporción de casos más satisfechos con el sistema antiguo que con el nuevo es menor del 1%.
- Cuando se dice que el nuevo sistema es mejor se tiene una probabilidad de error menor de 0,01.

Prueba de significación de una media

La aplicación del mecanismo anterior a una variable continua en la que se desea contrastar una hipótesis sobre su media es muy similar.

Ejemplo 7.3

Por ejemplo, se quiere poner a prueba si la media μ de la respuesta Y es una cierta media μ_H especificada en la hipótesis H .

Si la media μ de la población origen de la muestra es cierta media μ_H preespecificada, se escribe: $H : \mu = \mu_H$

Conocida σ , ya se sabe que \bar{y} será normal cuando o bien n sea grande ($n \geq 30$) o bien Y sea normal. Por tanto, bajo $H : \hat{Z} = (\bar{y} - \mu_H) / (\sigma/\sqrt{n}) \rightarrow N(0,1)$.

Y , como antes, puede calcularse el valor de P y creer H inverosímil si P es pequeña.

Nota técnica



$$P = P[|z| > |\hat{Z}|] = P[|z| > |(\bar{y} - \mu_H) / (\sigma/\sqrt{n})|]$$

Ejemplo 7.4



¿Recuerda el ejemplo 6.2, de la ASU para demostrar que las gasolineras estaban poniendo menos gasolina de la que cobraban? Se resolvió con un IC, pero ¿se puede demostrar que están timando? En una muestra aleatoria de 100 envases, aceptando $\sigma = 10$ ml, se debe tomar una decisión sobre si $\mu = 1.000$, habiendo observado una media $\bar{y} = 997$ ml.

i) Variable: contenido real en envases de 1.000 ml

ii) Estimador: promedio muestral \bar{y}

iii) H: $\mu_H = 1.000$ ml

iv) Se usará el estadístico $\hat{Z} = (\bar{y} - \mu_H) / (\sigma/\sqrt{n})$ que bajo H tiene una distribución Normal: $\hat{Z} \rightarrow N(0,1)$ siempre que (premisas) o bien la muestra sea grande ($n \geq 30$) o bien Y sea normal

v) Cálculo del estadístico: $\hat{Z} = (997 - 1.000) / (10/\sqrt{100}) = -3$
 $P = \text{Prob}[|z| > 3] < 0,0027$ (valor de Excel)

$P = 0,0027$, $\mu = 1.000$ ml es muy inverosímil

vi) Conclusión práctica: es poco verosímil que se esté dispensando la cantidad especificada.

vii) IC_{95%}: $\bar{y} \pm Z_{\alpha/2} \sigma/\sqrt{n} = 997 \pm 1,96 \cdot 10 / \sqrt{100} \approx 997 \pm 2 = [995, 999]$

La auténtica media μ de cantidad servida se encuentra entre 995 y 999 ml. Nos están timando, aunque a nivel individual, la cantidad es pequeña. La pequeña amplitud del IC_{95%} muestra que se dispone de mucha información.

Ejercicio 7.6



Los CD4 de una población de pacientes con sida tienen una desviación típica de 100. Se desea superar una media de 500 unidades. Se aplica una terapia experimental a una muestra de 36 casos de esta población y se obtiene una media muestral de 505. ¿Se puede afirmar que la media poblacional está por encima de 500?

Si, como es más usual, la varianza poblacional σ^2 es desconocida, se recurre a su estimador muestral S^2 y a la distribución t de Student. En este caso se usa el estadístico: $\hat{t} = (\bar{y} - \mu_0) / (S/\sqrt{n})$, que sigue una distribución t de Student con $n - 1$ grados de libertad, siendo «n» el número de casos.

Ejemplo 7.5

En 9 voluntarios sanos se ha estudiado la diferencia d entre los tiempos de respuesta a un estímulo visual y auditivo, habiéndose observado, $\bar{d} = 6,71$ y $S = 6,0$. Asumiendo que $d \rightarrow N$, ¿se puede aceptar que $E(d) = \mu = 0$, lo que implica que la respuesta a ambos estímulos es idéntica?

Solución:

i) Variable: diferencia entre el tiempo de respuesta a los estímulos visual y auditivo

ii) estimador: media de las diferencias \bar{d}

iii) $H: E(d) = \mu_H = 0$

iv) estadístico referencia $\hat{t} = (\bar{d} - \mu_H) / (S/\sqrt{n})$

Distribución bajo H : $\hat{t} \rightarrow t_{n-1} = t_8$

Premisa: d normal

v) Cálculos $\hat{t} = (6,71 - 0) / (6/\sqrt{9}) = 3,355$

$P = \text{Prob}[|t| > |3,355|] = 0,01$

vi) Como $P = 0,01$; $H: \mu_H = 0$ es poco inverosímil

Conclusión práctica: la igualdad entre ambos estímulos es dudosa.

vii) $IC_{95\%}: \bar{d} \pm t_{n-1, \alpha/2} S/\sqrt{n} = 6,71 \pm 2,306 \cdot 6/\sqrt{9} \approx 6,71 \pm 4,612 \approx [2,10, 11,32]$

La auténtica diferencia entre la respuesta media a ambos estímulos se encuentra entre 2,10 y 11,32.

Ejercicio 7.7

La satisfacción de los usuarios se mide por una escala que puntúa entre 0 y 100 cuya distribución se asemeja razonablemente a la normal. El objetivo de un servicio sanitario es conseguir satisfacciones por encima de 70. En una muestra al azar de 16 usuarios se ha observado una media de 79 y una desviación típica de 12. ¿Se puede afirmar que la media poblacional está por encima de 70?

Los estadísticos « \hat{z} » y « \hat{t} » como cociente señal / ruido

Los estadísticos \hat{z} y \hat{t} tienen una estructura muy similar. El numerador representa la distancia entre el valor de la muestra y el parámetro μ de la población. Y el denominador informa del error típico de \bar{y} , ya que como μ es un parámetro de la población (forma parte de la pregunta), no tiene error aleatorio de muestreo.

Ejemplo 7.4 (Cont.)

En el **ejemplo 6.2** del control de calidad de las gasolineras, si se desea saber si el surtidor cumple con las especificaciones (μ), este numerador representa la señal que proporciona la muestra: cuánto se desvía o distancia de la media especificada en la hipótesis. Se ha observado un valor de -3 .

Ejemplo

Por otro lado, la oscilación de \bar{y} explicable por el muestreo aleatorio puede cuantificarse en $\sigma_{\bar{y}} = \sigma / \sqrt{n} = 1$. Por tanto, el cociente «señal/ruido» vale -3 , lo que indica que la señal observada es negativa y 3 veces superior al error aleatorio medido por el error típico.

Recuerde

Tanto el estadístico \hat{z} como \hat{t} pueden ser vistos como un cociente señal / ruido.

Aun así, conviene reportar el valor exacto de P .

Pruebas de significación e intervalos de confianza

Digamos otra vez que las pruebas de significación (PS) y los intervalos de confianza (IC) son dos herramientas de inferencia: ambas permiten pasar de la muestra a la población. Mientras la prueba de significación hace una pregunta concreta o «cerrada» sobre el valor del parámetro en la población (¿es $\pi = 0,5$?), el intervalo de confianza hace una pregunta «abierta», (¿cuál es el valor de π ?). Se podría argumentar que el intervalo es una herramienta positiva que dice cuáles son los posibles valores del parámetro compatibles con la muestra observada, mientras que la prueba de significación es una herramienta negativa.

Ejemplo 7.6

Recuperemos el ejemplo de las 52 caras en 100 lanzamientos de una moneda. El intervalo de confianza del auténtico valor de la probabilidad de cara era:

$$IC_{0,95} \pi = p \pm Z_{\alpha/2} \sqrt{p(1-p)/n} = 0,52 \pm 1,96 \sqrt{[0,52 \cdot 0,48 / 100]} \approx 0,52 \pm 0,10 = [0,42, 0,62]$$

Se cree, con una confianza del 95%, que esta moneda tiene una probabilidad de cara, π , situada entre el 42 y el 62%. Este resultado coincide con el de la prueba de significación que, con una $P = 0,69$, no permitía rechazar la hipótesis de que la probabilidad de cara era 0,5.

En el caso de observar 63 caras el IC es:

$$IC_{0,95} \pi = p \pm Z_{\alpha/2} \sqrt{p(1-p)/n} = 0,63 \pm 1,96 \sqrt{[0,63 \cdot 0,37 / 100]} \approx 0,63 \pm 0,095 = [0,535, 0,725]$$

Por lo que ahora se cree, con una confianza del 95%, que esta probabilidad de cara, π , es alguno de los valores comprendidos entre el 53,5 y el 72,5%. Dado que este intervalo excluye el valor 0,5, se llega a la misma conclusión que con la prueba de significación, que había arrojado un valor de $P = 0,01$.

Las conclusiones de la prueba de significación y del intervalo de confianza coinciden.

Se pueden utilizar los intervalos de confianza para poner a prueba la hipótesis de interés, ya que los valores del parámetro excluidos del IC generarían PS con valores de P poco verosímiles.

Ejercicio 7.8



En 100 pacientes con sida el intervalo de confianza al 95% de la media μ del recuento de CD4 va de 375 a 500. Si se plantearan las dos pruebas de significación siguientes con $\alpha = 0,05$: (A) $H_A: \mu = 400$; (B) $H_B: \mu = 350$ las conclusiones serían:

- nada se opone a aceptar ambas H ;
- se rechazan ambas H ;
- nada se opone a aceptar H_A y se rechaza H_B ;
- se rechaza H_A y nada se opone a aceptar H_B .

El intervalo de confianza completa la información de la prueba de significación y ayuda a interpretarla, ya que informa sobre los valores plausibles del parámetro.

Nota técnica



En el caso de rechazar una hipótesis H (p. ej., $\pi = 0,5$), la conclusión formal de la prueba de significación sería que H es inverosímil sin decantarse hacia ninguno de los dos lados. Pero, a nivel práctico, el intervalo de confianza permite conocer, no sólo el lado, sino también los valores razonables del parámetro.

Cuando no se rechaza H , el intervalo de confianza permite distinguir entre poca (IC muy amplio) y mucha información (IC estrecho).

Recuerde



Utilice siempre IC. Incluso si el objetivo principal es la prueba de significación, conviene acompañar sus resultados con un IC. Si se rechaza H , porque se dice dónde se cree que está el parámetro. Si no se rechaza H , porque se cuantifica la información de que se dispone.

Comentario



Cuando no se puede rechazar la hipótesis, la prueba de significación concluye: «No es inverosímil: no hay evidencia en contra de H ». Pero ello puede ser, bien por falta de evidencia para establecer algo existente (¿muestra pequeña, diseño deficiente, análisis pobre, etc.?), o bien porque realmente no hay nada que ver.

Recuerde



*En PS, ausencia de pruebas no es prueba de ausencia.
En PS, ausencia de evidencia no es evidencia de ausencia.*

Lectura



Las recomendaciones para los autores de revistas biomédicas (9) anteponen el uso de IC al de PS: «A pesar de que los valores de P se pueden ofrecer junto a los intervalos de confianza, los resultados no deben recoger únicamente los valores de P» (CONSORT, ítem 17).

Pruebas de significación unilaterales y bilaterales

Hasta el momento, sólo se han planteado pruebas **bilaterales** (también llamadas de dos colas), como el ejemplo de la moneda, que se consideraba defectuosa tanto si salían demasiadas caras como demasiado pocas. En consecuencia, la región crítica de la hipótesis se sitúa a ambos lados (fig. 7-4):

$$H : \pi = 0,5$$

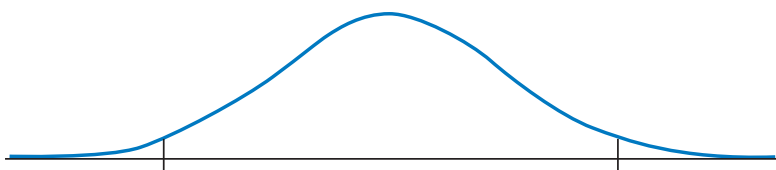


Figura 7-4 Las pruebas bilaterales contemplan los casos en contra de H en ambas colas.

Pero se pueden plantear también pruebas de una sola cola. En el ejemplo de la gasolinera, en el que se quería detectar si había tимо, ¿qué se puede concluir si la media observada se situaba por encima de la media teórica? ¿Que regalan gasolina? En esta situación, tiene más sentido una prueba **unilateral por la izquierda** (fig. 7-5):

$$H : \mu \geq \mu_0$$

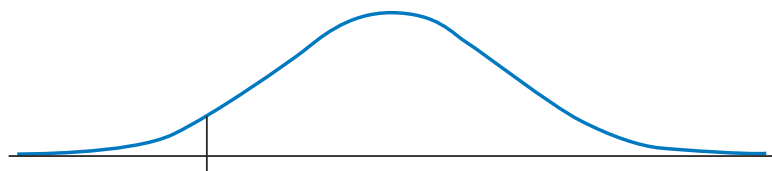


Figura 7-5 Las pruebas unilaterales por la izquierda contemplan los casos en contra de H en el lado izquierdo.

Y, de forma simétrica, si se estudia cómo aumenta la respuesta al aumentar la dosis, podría tener más sentido una prueba **unilateral por la derecha** (fig. 7-6):

$$H : \mu \leq \mu_0$$

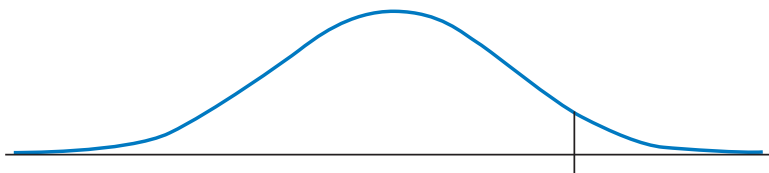


Figura 7-6 Las pruebas unilaterales por la derecha contemplan los casos en contra de H en el lado derecho.

Recuerde



En el caso de pruebas unilaterales, debe considerarse sólo una cola.

Resaltemos dos aspectos relevantes:

1) Al concentrar todo el nivel de significación en un lado, se hace algo mayor la región crítica de esa cola, por lo que una hipótesis unilateral o bilateral, ¡puede cambiar las conclusiones!

2) El signo igual (acompañado, ahora, por el desigual) sigue figurando en H .

Recuerde



La hipótesis H es el punto de salida (no el de llegada), y condiciona todo el experimento: la hipótesis H debe establecerse previamente a la recogida de datos.

Ejercicio 7.9



Repita los ejercicios 7.6 y 7.7 bajo un planteamiento unilateral.

Ejercicio 7.10

Se desea resolver la prueba $H: \mu \leq 0$ mediante un estadístico que sigue una distribución normal $(0,1)$. El resultado de la prueba ofrece $z = -2$, por lo que se concluye (elija una):

- que la media poblacional es 0;
- que la media poblacional es mayor que 0 ($P = 0,0227$);
- que la media poblacional es menor que 0 ($P = 0,0227$);
- hay una probabilidad del 97,37% de que la media poblacional sea 0;
- nada se opone a aceptar la hipótesis de que la media es igual o inferior a 0.

Decisión: contraste de hipótesis

El contraste de hipótesis es un instrumento para tomar una decisión manteniendo controlados los riesgos de error.

Ejemplo 7.7



(Prestado de un examen para informáticos de la profesora Monique Becue.) Para creer si cierto «garabato» es un 8, un programa de reconocimiento de patrones mide la curvatura izquierda (Y), cuya distribución tiene una media de 12 u si se trata de un «8», y una media superior si se trata de una «B». Se sabe que la distribución de Y es normal y que $\sigma = 3$ u. Si se está dispuesto a aceptar que un 5% de ochos (8) sean reconocidos como «bes» (B), ¿a partir de qué valor se dirá que se trata de una «B»? Es decir: ¿cuál es el límite de la región crítica?

i) $H : \mu = 12$ u (se trata de un 8)

ii) Se usará el estadístico $\bar{Z} = (\bar{y} - \mu_H) / (\sigma/\sqrt{n})$

Distribución bajo $H : \bar{Z} \rightarrow N(0,1)$

Premisas: dado que $n = 1$, Y debe ser normal

iii) Regla rechazo. Con $P = 0,05$, se rechazará H si $\hat{Z} > Z_\alpha = 1,645$.

iv) Cálculo del límite: $\mu_0 + Z_\alpha (\sigma/\sqrt{n}) = 12 + 1,645 \cdot 3 = 16,935$

Si $y > 16,935$ se «cree» que no se trata de un 8.

Supóngase ahora que se conoce que la distribución de las «B» es $N(21,3)$. Aceptando como límite de decisión $y = 16,935$, se desea calcular la probabilidad de que una B sea reconocida como un 8 (fig. 7-7). Ahora se dispone de dos situaciones hipotéticas, entre las que se debe escoger.

$H_0 : \mu = 12$ (se trata de un 8)

$H_1 : \mu = 21$ (se trata de una B)

Prob [$y \leq 16,935$ condicionado a $Y \rightarrow N(21,3)$] =

= $P [Z \leq (16,935 - 21) / 3] =$

= $P (Z \leq -1,355) \approx 0,0885$

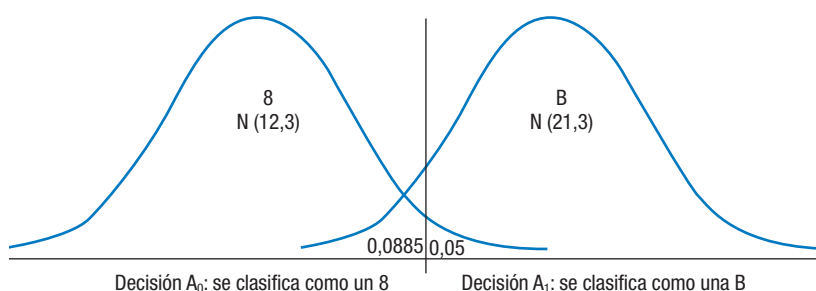


Figura 7-7 Si el valor observado supera el límite 16,935 se clasifica el garabato como B (A_1) y en caso contrario como 8 (A_0).

Comentario

En lo que sigue emplearemos A_0 y A_1 (acción 0, acción 1) para resaltar la acción que implica el contraste de hipótesis. Sea A_0 «conservadora» y A_1 «innovadora». Para tomar la acción A_1 hace falta rechazar H_0 .

Definición

Un contraste de hipótesis permite elegir entre dos acciones alternativas.

Se han identificado dos actuaciones erróneas y se han cuantificado los riesgos respectivos:

$$P(\text{concluir } B \mid \text{realidad } 8) = 0,05$$

$$P(\text{concluir } 8 \mid \text{realidad } B) = 0,088$$

Un organismo científico, como la Sociedad Española de Neurología, la revista *Medicina Clínica* o la Cochrane Collaboration, está interesado en lo que científicamente se sabe y, por tanto, en realizar intervalos de confianza o pruebas de significación. En cambio, un órgano ejecutivo, como una agencia reguladora del medicamento o un comité que elabora protocolos, debe proponer decisiones, acciones concretas.

Ejemplo 7.8

Fisher (41) y Hill (42) mantuvieron posiciones distintas en cuanto a la evidencia disponible sobre los efectos del tabaco. Sea cual sea esta evidencia, a un responsable de Salud Pública, lo que le concierne es, a la luz de dicha información, cuál debe ser su actuación. Greenland (42b) recuerda que un organismo de Salud Pública debe actuar y debe, por tanto, tomar decisiones: ante humo en un bosque, la acción pertinente es enviar bomberos, no científicos para averiguar si debajo del humo hay fuego. A nivel personal, por ejemplo, un fumador debe valorar las consecuencias de los dos «errores» posibles: *i)* que decida seguir fumando, pero tenga razón Hill y él mismo sea de la proporción de casos que desarrollan el cáncer hacia los 50 años; o *ii)* que decida no fumar, pero tenga razón Fisher y no se «ahorre» dicha enfermedad. Cada uno debe valorar qué consecuencias tiene cada posible situación.

Historieta

Supongamos que quien debe decidir si fuma, o no, se plantea minimizar su sentimiento de estupidez. Para eso debe pensar con cuál de los dos posibles errores se sentirá más estúpido: i) ¡Qué lástima! podría haber vivido 30 años más; ii) ¡Qué lástima!, podría haber fumado.

Límites de significación

El límite de significación a partir del cual se rechaza H tiene un equivalente en la escala de los estadísticos, \hat{Z} o \hat{t} . En la escala Z , los límites que corresponden a $P = 0,05$ son $-1,96$ y $+1,96$. En la t de Student, dependerá de los grados de libertad.

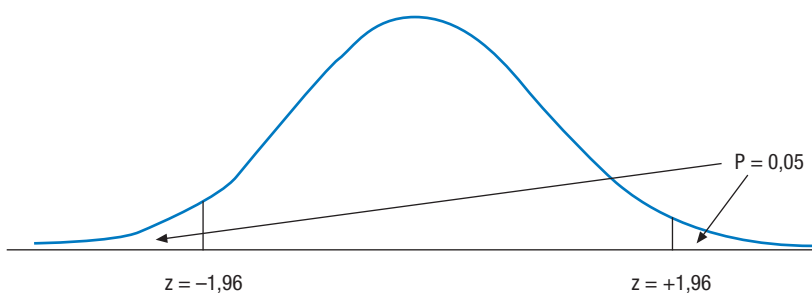


Figura 7-8 Es equivalente preguntarse si $P < 0,05$ o si \hat{Z} es mayor que 1,96 o menor que $-1,96$.

Ejercicio 7.11

¿Cómo se habría realizado la comparación con Z en los ejercicios 7.3, 7.4 y 7.6? ¿Y con t en el 7.7?

Recuerde

La prueba de significación contesta la pregunta «qué creo» y el contraste de hipótesis, «qué hago».

Errores tipo I y II. Riesgos α y β

En el contraste de hipótesis se definen dos tipos de errores.

Definición

El error de primera especie o tipo I consiste en tomar la acción alternativa (A_1) cuando era cierta H_0 .
Tomar A_1 | es cierta H_0

Ejemplo 7.9

Concluir que es una B cuando en realidad es un 8 es un error tipo I.

Definición

El error de *segunda especie* o *tipo II* consiste en tomar la acción nula (A_0) cuando es cierta H_1
Tomar A_0 | es cierta H_1

Ejemplo 7.9 (Cont.)

Concluir que es un 8 cuando en realidad es una B es un error tipo II.

Definición

Las probabilidades correspondientes de cometer errores de primera y de segunda especie reciben el nombre de *riesgos* α y β :
Riesgo $\alpha = P(\text{Tomar } A_1 \mid \text{es cierta } H_0)$
Riesgo $\beta = P(\text{Tomar } A_0 \mid \text{es cierta } H_1)$

Ejemplo 7.10

Riesgo $\alpha = P(\text{Decidir es una B} \mid \text{en realidad es un 8})$
Riesgo $\beta = P(\text{Decidir es un 8} \mid \text{en realidad es una B})$
De esta manera, α representa la proporción de 8 que serán identificados como B y β su recíproco.

Ejemplo 7.11

Un laboratorio farmacéutico propone a una agencia reguladora del medicamento un ensayo clínico para contrastar H_0 (misma eficacia que control) frente H_1 (eficacia mayor que el control = Δ). Si se rechaza H_0 , el fármaco se pondrá en el mercado (A_1). El riesgo α sería la proporción de medicamentos iguales que el control (H_0) que son finalmente puestos en el mercado (A_1). A su vez, el riesgo β , es la proporción de medicamentos que superan el control en un valor Δ (H_1) que no llegan al mercado (A_0).

Ejercicio 7.12



Un proveedor entregaba un reactivo con un tiempo de reacción medio de 100 ms y desviación típica de 10 ms. Ahora, ofrece uno nuevo con parámetros $m = 50$ ms y $s = 5$ ms. Sea H_0 : $m = 100$ ms y $s = 10$ ms (reactivo viejo); H_1 : $m = 50$ ms y $s = 5$ ms (reactivo nuevo); A_0 : rechazar el lote y A_1 : aceptar el lote. El riesgo α de cometer un error de primera especie es:

- la probabilidad de que el reactivo sea nuevo;
- la probabilidad de aceptar un lote (A_1) de reactivos viejos (H_0) [ante un reactivo viejo, la probabilidad de decidir que es de los nuevos];
- la probabilidad de rechazar un lote (A_0) de reactivos nuevos (H_1) [ante un reactivo nuevo, la probabilidad de decidir que es de los viejos];
- todas son falsas.

Ejercicio 7.13

En un contraste de hipótesis, si H_0 es cierta, es posible (elija una):

- cometer dos errores, el de tipo I y el de tipo II;
- sólo se puede producir el de tipo I;
- sólo se puede producir el de tipo II;
- ninguno, ya que H_0 es cierta.

Definición



La potencia (tabla 7-1) de un estudio es $1 - \beta$ o probabilidad de decidir A_1 cuando es cierta H_1 :

$$\text{Potencia} = 1 - \beta = P(\text{Decidir } A_1 \mid \text{es cierta } H_1)$$

Tipos de errores y riesgos		Decisión	
		A_0	A_1
Realidad	H_0		Tipo I (riesgo α)
	H_1	Tipo II (riesgo β)	Potencia = $1 - \beta$

Tabla 7-1 Resumen de tipos de errores y riesgos

Intervalos de confianza, pruebas de significación y contraste de hipótesis

En un intervalo de confianza, IC, el nivel de confianza α se decide a priori. En un contraste de hipótesis, CH, también, y se opta por aquel diseño y estadístico que minimiza β , que también se establece a priori. Así, en el entorno de IC y CH, lo único que tiene valor y debe, por tanto, ser reportado son los valores de α y β decididos a priori. En cambio, en la prueba de significación, PS, el valor de P es un resultado ob-

tenido al final del experimento y el nivel de evidencia que aporta en contra de H sería diferente ante un valor de $P = 0,023$ o de $P < 0,001$, por lo que se recomienda reportar el valor de P exacto.

Resumen



*En IC se debe informar del valor de α fijado a priori.
En PS se debe reportar el valor exacto obtenido de P .
En CH se debe informar de los valores de α y β fijados a priori.
La misma concordancia en el cálculo que existe entre IC y PS se aplica también a CH. En cambio, los resultados de cada técnica deben interpretarse de acuerdo con sus objetivos.*

Resumen



*IC, PS y CH difieren en objetivos:
IC, estimar valores del parámetro
PS, aportar evidencia en contra de H
CH, decidir entre A_0 y A_1 minimizando los riesgos α y β
Pero coinciden en su mecánica:
IC $(1 - \alpha)$: $\mu \in (\bar{y} \pm 1,96 \cdot \sigma/\sqrt{n})$
PS (P): considerar inverosímil H si $\bar{y} \notin (\mu_H \pm 1,96 \cdot \sigma/\sqrt{n})$
CH (α, β) : decidir A_0 si $\bar{y} \in (\mu_0 \pm 1,96 \cdot \sigma/\sqrt{n})$*

Nota técnica



En algunas situaciones, la estimación de la varianza del estimador no es la misma bajo H , H_0 , H_1 o en la muestra observada, por lo que IC, PS y CH no coincidirán plenamente. Por ejemplo, en el caso de π y p la amplitud de los intervalos puede diferir:

PS (p): aceptar si $p \in \pi_H \pm 1,96 \sqrt{[\pi_H (1 - \pi_H) / n]}$

IC $1 - \alpha$: $\pi \in p \pm 1,96 \sqrt{[p(1 - p) / n]}$

En el modelo lineal (comparación medias, regresión, etc.) sí que coinciden.

Interpretación errónea de P y α

Nota técnica



Es importante remarcar que en la tabla 7-1 los riesgos α y β representan probabilidades condicionadas a la fila, no a la columna. Es decir, la probabilidad de una conclusión dada la realidad de una hipótesis. Nótese que las filas representan valores del parámetro, que es una constante, mientras que las columnas representan zonas en las que se sitúa el estadístico, que sí que es una variable aleatoria.

Así, en un contraste de hipótesis (H_0 frente a H_1) para tomar una decisión (A_0 frente a A_1), α y β representan la proporción o frecuencia de decisiones erróneas a largo plazo. En el ejemplo 7.11, α es la proporción de fármacos iguales que el control (H_0) que a largo plazo son puestos en

Nota técnica



el mercado (A_1) y β , la proporción de fármacos que superan el control en un valor Δ (H_1) que no son puestos en el mercado (A_0).
Nótese que en una prueba de significación, P (y su máximo aceptado, 0,05) indica el nivel de evidencia en contra de H , mientras que en un contraste de hipótesis α y β indican la frecuencia de decisiones erróneas.

Recuerde



P en la prueba de significación es medida de información empírica («evidencia») en contra de H , mientras que α y β en el contraste de hipótesis cuantifican la frecuencia de decisiones erróneas.

Nota técnica



En la tabla 7-1, α se lee en la primera fila como la proporción de veces en que se toma la decisión A_1 del total de ocasiones en que se parte de H_0 . Otra cosa muy distinta sería la lectura en la segunda columna de la proporción de casos provenientes de H_0 del total de ocasiones en que se ha tomado la acción A_1 . Si en el ejemplo 7-11 $\alpha = 0,05$ y $\beta = 0,10$, ambos bilaterales, un 5% (primera fila) de fármacos no eficaces (H_0) llegarán al mercado (A_1) y un 10% (segunda fila) de fármacos de eficacia Δ (H_1) no llegarán al mercado (A_0), pero la proporción (segunda columna) de fármacos no eficaces entre los que llegan al mercado es desconocida.

La tabla 7-2 expone los términos que pueden emplearse para informar el resultado de una PS o un CH.

PRUEBA DE SIGNIFICACIÓN		
Si el valor de P es...	Grande (p. ej. 0,634)	Pequeño (p. ej. 0,0001)
H es...	verosímil	inverosímil
La diferencia...	es explicable por...	no es explicable por...
La diferencia...	no es estadísticamente significativa	si es estadísticamente significativa
A nivel práctico...	no hemos logrado demostrar que la moneda está cargada	creemos que la moneda está cargada
CONTRASTE DE HIPÓTESIS		
Si el estadístico se sitúa en...	Región de aceptación	Región crítica
	Se acepta H_0	Se rechaza H_0
	Se toma la acción A_0	Se toma la acción H_1

Tabla 7-2 La prueba de significación y el contraste de hipótesis en palabras

Recuerde

Ni el riesgo α ni el valor de P pueden resumirse por «la probabilidad que tengo de haberme equivocado».

Ejercicio 7.14

¿Cuál o cuáles son correctas?

- a) El valor P es la probabilidad de equivocarse.
- b) El valor P es la probabilidad de equivocarse al rechazar H .
- c) El valor P es la probabilidad de equivocarse al aceptar H .
- d) El valor P es la probabilidad de observar el resultado actual (o más discrepante) si fuera cierta H .
- e) El riesgo α es la probabilidad de equivocarse;
- f) El riesgo α es la probabilidad de equivocarse al rechazar H .
- g) El riesgo α es la probabilidad de equivocarse al aceptar H .
- h) El riesgo α es la frecuencia esperada de ocasiones en las que siendo cierta H_0 tomaremos la decisión (errónea) A_1 .
- i) El riesgo β es la probabilidad de equivocarse.
- j) El riesgo β es la probabilidad de equivocarse al rechazar H .
- k) El riesgo β es la probabilidad de equivocarse al aceptar H .
- l) El riesgo β es la frecuencia esperada de ocasiones en las que siendo cierta H_1 tomaremos la decisión (errónea) A_0 .

Ejemplo 7.12

La celebración final de carrera ha sido magnífica. A las 5 de la mañana los amigos se despiden, pero uno de ellos decide seguir la juerga y le pide al taxista que le lleve a una buena partida de póquer. Tras pasar los controles típicos, que su amigo creía cosa de película, consigue entrar en un 5.º piso de la calle Enrique Granados, donde se sienta a una mesa y empieza a perder dinero. Sus rivales no paran de sacar magníficas jugadas. Tanto, que él calcula que, asumiendo que no hacen trampas, la probabilidad de esos resultados (o incluso mejores) es de tan sólo una entre un millón. ¿Qué hace? Por supuesto, deja de jugar. El valor de $P = 0,000001$ le permite rechazar la hipótesis de que no le hacen trampas.

Ejemplo 7.13

En la celebración de las Navidades, un joven investigador vuelve del hospital Mount Sinai para visitar a su familia y acaban jugando al póquer con idénticos resultados que en el ejemplo anterior. A pesar de que este investigador calcu-

Ejemplo 7.13 (Cont.)

la el mismo nivel de significación anterior (asumiendo que no hacen trampas, esos resultados o mejores sólo ocurren 1 vez por millón), sigue jugando confiado, ya que no se plantea la posibilidad alternativa, de que su familia le haga trampas. Por lo que dice, «caramba, qué mala suerte tengo hoy».

Nota técnica

La estadística bayesiana lamenta que la solución de los dos ejemplos anteriores no tenga en cuenta toda la información contenida en el enunciado. Antes de empezar a jugar, el primer titulado ya podía sospechar que le harían trampas, pero no el segundo. Para poder calcular, a partir de los resultados muestrales, la probabilidad de que una hipótesis sea cierta, es preciso recurrir a una formalización del conocimiento científico previo: antes de los datos que actualmente se están analizando, ¿qué se sabía sobre este tema?, ¿qué se sabía sobre el valor del parámetro? Si se acepta representar el nivel de incertidumbre previa en forma de probabilidades sobre los diferentes valores del parámetro, ya se tienen los elementos necesarios para actualizar la información científica mediante el teorema de Bayes.

Ejercicio de Navegación

Referencias críticas sobre el abuso de las pruebas de significación, así como enlaces a paginas web aplicadas, y un *applet* muy instructivo, pueden encontrarse en:
<http://www.stat.duke.edu/~berger/p-values.html>

Recuerde

La prueba de significación, el contraste de hipótesis y el intervalo de confianza se concentran en la información aportada por los datos actuales, pero no la «suman» a la información previa.

Sólo el contraste de hipótesis permite «aceptar H_0 »

La PS no especifica ninguna hipótesis alternativa y, por tanto, no tiene definida ninguna medida análoga al riesgo β . En consecuencia, la PS no tiene ningún argumento para defender la credibilidad de H , su única hipótesis.

Nota técnica

El riesgo β puede delimitarse cuando el contraste de hipótesis tiene, como en el ejemplo del 8 y la B, la forma:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases}$$

Pero si la prueba de significación es de la forma:

$$H : \mu = \mu_H$$

Entonces la definición de una medida análoga al riesgo β bajo todos los posibles $\mu \neq \mu_H$ llevaría al cálculo de infinitos riesgos β , tantos como valores posibles de μ . Pero para valores próximos a μ_H ese riesgo tiende hacia $1 - \alpha$ (fig. 7-9).

La prueba de significación sólo define una hipótesis H . Este planteamiento «asimétrico» conduce a una conclusión asimétrica: si el valor de P es pequeño, se considera H inverosímil. En cambio, si P es grande, «nada se opone a aceptar H ».

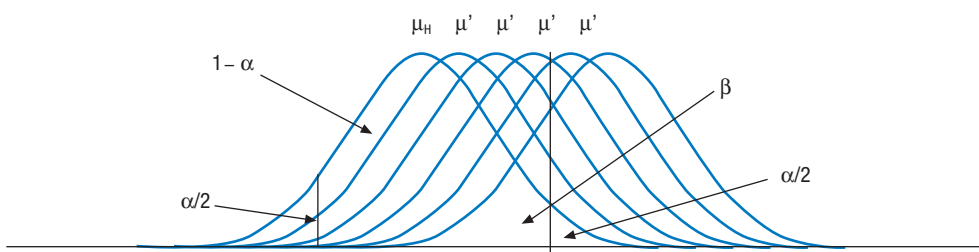


Figura 7-9 Si no hay hipótesis alternativa cerrada, el riesgo beta no está acotado.

Recuerde

En PS, «ausencia de pruebas» no es «prueba de ausencia». En PS diga «no se han detectado diferencias» en lugar de «no existen diferencias». El CH, al tener acotados α y β , permite tomar ambas decisiones.

Ejercicio 7.15

El laboratorio Yotambién S.L., para demostrar que su genérico es tan eficaz como el de la compañía Los primeros S.A., realiza un ensayo en el que compara ambos productos. Supóngase que obtiene un valor de $P = 0,23$, ¿puede concluir que ambos productos tienen la misma eficacia?

Conviene ir con mucho cuidado con las palabras que se utilizan para explicar las conclusiones de una prueba de significación. La tabla 7-2 resume algunas de las más habituales. Nótese la asimetría de la conclusión a la que se llega en ambas regiones: mientras en la zona crítica se afirma que se rechaza H («se ha demostrado la culpabilidad del acusado»), en la zona de aceptación no hay afirmaciones rotundas («absuelto por falta de pruebas»).

Ejercicio 7.16



¿Es alguna/s de las siguientes falsa?:

- En la PS se buscan evidencias en contra de H .
- El CH permite tomar ambas decisiones, A_0 y A_1 .
- Tanto P como α cuantifican áreas de las distribuciones de probabilidad, pero miden aspectos distintos.
- En la PS debe reportarse el valor exacto del valor de P .
- En el CH debe reportarse el valor previo α , usualmente, 0,05.
- En la PS, si $P > 0,05$, nada se opone a aceptar H .
- Una ventaja del CH es que permite decidir A_0 y A_1 .
- Una ventaja del CH es que cuantifica β .

Ejercicio 7.17

La prueba de significación es conservadora en el sentido de que se declara ??? la hipótesis hasta que no haya clara evidencia en su contra:

- ??? = cierta
- ??? = falsa
- La prueba de significación no es conservadora.
- Todas son incorrectas.

Ejercicio 7.18

En un estudio para comparar dos tratamientos (H_0 : son iguales) $P = 0,341$, ¿cuál o cuáles son ciertas?

- Nada se opone a aceptar H_0
- No existen diferencias
- No se han detectado diferencias
- La probabilidad de que sean diferentes es 0,341.

Lectura



Karl Popper (43) ha contribuido a incorporar los avances estadísticos en la epistemología o metodología científica. De acuerdo con esta asimetría de la conclusión de una prueba de significación, afirmó que lo único que se puede hacer con una teoría científica es ponerla a prueba y rechazarla en el caso de que encontremos pruebas en su contra, pero que nunca se podrá demostrar que sea cierta y constituya la última palabra de la ciencia en ese punto. Así, Popper dice que el criterio para establecer el estatus científico de una teoría es su refutabilidad o su testabilidad: «Para ser colocados en el rango de científicos, los enunciados o sistemas de enunciados deben ser susceptibles de entrar en conflicto con observaciones posibles», lo que es conocido como problema de la demarcación. Así, una teoría científica es más fuerte cuando más falseable es, es decir, cuanto más fácilmente podría demostrarse su falsedad (caso de ser falsa).

Interpretación del contraste de hipótesis

Desde un punto de vista formal, disponer de dos hipótesis simples, cada una con un único valor, permite definir muchas propiedades interesantes para escoger el «mejor» estadístico. Los libros clásicos de estadística matemática exponen la teoría desarrollada por Pearson y Newman sobre el contraste de dos hipótesis simples. Fisher se centra en la inferencia sobre una hipótesis, por lo que sólo puede cuantificar P y sólo puede rechazar H en lo que él llama prueba de significación.

Recuerde



La prueba de significación es un instrumento de inferencia; si la P es pequeña, Fisher recomienda modificar nuestras opiniones sobre la veracidad de H .

El contraste de hipótesis es un instrumento de decisión que permite acotar los riesgos de tomar acciones erróneas.

En la prueba de significación hay que reportar el valor exacto de P , por ejemplo, 0,0015; mientras que en el contraste de hipótesis hay que informar sobre las frecuencias α y β de errores a largo plazo, por ejemplo, $\alpha = 0,05$ y $\beta = 0,020$.

Historieta



Si quiere que los estadísticos no influyan en la reunión, póngalos en número par.

Detrás de las dos decisiones diferentes en las partidas de póquer de los ejemplos 7.12 y 7.13, la del garito y la de Navidad, también hay dos consecuencias muy diferentes. Levantarse de la partida de póquer del garito (sin ofender, claro) no debería conllevar consecuencias. Por su parte, seguir la partida con nuestra madre no conlleva pérdidas: incluso, en el caso de que sea cierto que hace trampas, «el dinero se queda en casa».

Ejemplo 7.14



Ya se ha dicho que Gosset era un estadístico que trabajaba en la cervecera Guinness, en su departamento de control de calidad, donde se planteaba la decisión de aceptar o rechazar una barrica de cerveza. Además de los riesgos α y β , debía considerar los costes por desechar una barrica correcta y por poner en el mercado una que no lo era.

Recuerde



El proceso de decisión, además de los riesgos de error, debe valorar también sus consecuencias, el coste que se paga por cada decisión errónea y el premio que se obtiene con las decisiones correctas.

Lectura



Uso del paracaídas para prevenir el fallecimiento y el traumatismo grave relacionados con la exposición a la fuerza de la gravedad: una revisión sistemática de los ensayos clínicos efectuados con asignación aleatoria y control [Smith et al. (44), publicado en el número bufo de Navidad]

Objetivos. Determinar la efectividad de los paracaídas para la prevención de los traumatismos graves relacionados con la exposición a la fuerza de la gravedad.

Diseño. Revisión sistemática de los ensayos clínicos efectuados con asignación aleatoria y control.

Fuentes de datos. Medline, Web o Science, Embase y las bases de datos de la Cochrane Library; sitios web apropiados y listas de bibliografía recogidas en Internet.

Selección de los estudios. Estudios en los que se demuestran los efectos del uso del paracaídas en situaciones de caída libre.

Criterio principal de valoración. Fallecimiento o traumatismo grave, definido como un traumatismo con una puntuación de gravedad > 15 .

Resultados. No ha sido posible identificar ningún ensayo clínico efectuado con asignación aleatoria y control relativo al uso del paracaídas en situaciones de caída libre.

Conclusiones. Al igual que ocurre con muchas otras intervenciones para prevenir problemas de salud, la efectividad del paracaídas no ha sido objeto de una evaluación rigurosa a través de ensayos clínicos realizados con asignación aleatoria. Los defensores de la medicina basada en la evidencia han criticado la adopción de intervenciones evaluadas únicamente a través de datos de observación. Consideramos que sería de gran utilidad que los defensores más radicales de la medicina basada en la evidencia diseñaran y llevaran a cabo un ensayo clínico sobre la efectividad del paracaídas que fuera: enmascarado, controlado con placebo y con diseño cross-over, en el que cada caso es sometido a las dos intervenciones en comparación.

Ejemplo 7.15



Es bien conocido que aunque un tratamiento puede haber demostrado un cierto efecto positivo en una variable de interés, sus costes pueden aconsejar antes otra intervención sanitaria más eficiente, en el sentido de que una misma «inversión» origine un mayor «retorno», valorado en términos de salud.

Ejemplo 7.16



La pregunta de quien desea optimizar el empleo de los recursos sanitarios es: la inversión de este euro, ¿dónde me genera un mayor retorno en términos de salud?

Nota técnica

El contraste de hipótesis es el primer instrumento de la teoría de la decisión, que constituye toda una rama de la estadística y es ampliamente utilizada en otras disciplinas, como por ejemplo, la economía, donde los «costes» y los «premios» son fácilmente expresables en una única escala. El diagnóstico y el tratamiento son dos ejemplos de acciones médicas que podrían beneficiarse de las aportaciones de la teoría de la decisión.

Ejercicio 7.19

¿Cuáles de las siguientes frases son de inferencia y cuáles de decisión?

- a) el riesgo es mayor en pacientes de tipo A;
- b) el riesgo disminuye a la mitad si se adoptan las medidas X;
- c) la obesidad abdominal es el componente de síndrome metabólico de mayor prevalencia en mujeres;
- d) el valor predictivo de la escala de Z implica que debería utilizarse en el futuro para clasificar a este tipo de enfermos;
- e) si hay dos o menos factores de riesgo presentes y la PAS ≥ 160 o la PAD < 100 (siendo PAS < 180 y PAD < 110), conviene intentar cambios en el estilo de vida durante varios meses y luego, si se mantiene, tratamiento farmacológico.

Puede ser razonable esperar que el efecto de una intervención sea el mismo en diferentes condiciones (país, entorno de atención al paciente, raza, etc.). Y también la capacidad predictiva de un indicador. Pero no es en absoluto razonable esperar que las consecuencias de una decisión se valoren igual en diferentes entornos. Por ejemplo, el «valor» del mismo coste de un medicamento puede diferir de un país a otro.

Recuerde

El proceso de decisión implica una valoración de las consecuencias, que tienen connotaciones locales y es más difícilmente extrapolable que la mera inferencia de conocimiento.

Ejercicio adicional

Encuentre un original científico reciente que, en su discusión, vaya de la interpretación de los resultados de inferencia a la decisión ulterior sin considerar formalmente el proceso de decisión, sus riesgos y sus consecuencias en el entorno en el que propone la acción.

Nota técnica

Una ventaja del planteamiento bayesiano es que, de forma natural, conduce desde la inferencia científica hasta la toma de decisiones práctica. Permite, pues, combinar la cotidiana toma de decisiones de un profesional con la adquisición científica de conocimiento. En otras palabras, hace transparente el paso de la investigación epidemiológica a las decisiones de salud pública o de la investigación farmacológica a las decisiones de las agencias de tecnología sanitaria.

Equivalencia

Usualmente interesa establecer «diferencias».

Ejemplo 7.17

Por ejemplo, «el riesgo de sida es mayor en toxicómanos por vía parenteral», o bien «el nuevo tratamiento es mejor que el clásico».

Pero también puede interesar establecer «equivalencia».

Definición

Un tratamiento es **equivalente** a otro si la diferencia de sus efectos no alcanza un cierto valor Δ que hace relevantes las consecuencias.

Ejemplo 7.18

Se desea establecer que: $\Delta_1 < \text{efecto} < \Delta_2$.
 Δ_1 y Δ_2 delimitan el intervalo de equivalencia.

cercanos a la igualdad, que no son relevantes desde el punto de vista práctico.

Recuerde

Ambos límites deben ser rechazados para poder establecer equivalencia.

Ejemplo 7.19

Se desea establecer, de forma simétrica, que $|\text{efecto}| < \Delta$.

El concepto de equivalencia es más amplio que el de la estricta igualdad, pues incluye también aquellos valores, cercanos a la igualdad, que no son relevantes desde el punto de vista práctico.

La figura 7-10 representa todos los posibles valores de la diferencia entre las dos medias de interés: $\mu_1 - \mu_2$.

Para establecer equivalencia se debe demostrar que las diferencias no alcanzan ni superan los dos límites especificados. Y ello puede hacerse mediante un intervalo de confianza que deberá quedar comprendido entre estos límites, lo que equivale a realizar dos pruebas que deberán rechazar ambos límites.

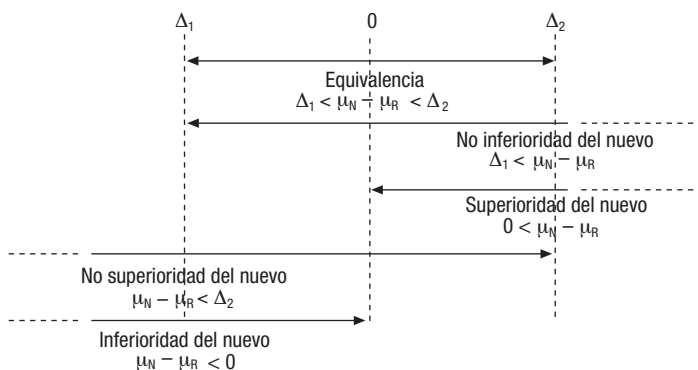


Figura 7-10 Definición de los conceptos de superioridad, equivalencia y no inferioridad.

Ejemplo 7.20



La figura 7-11 muestra 3 estudios en los que se concluiría equivalencia y 3 estudios en los que no.

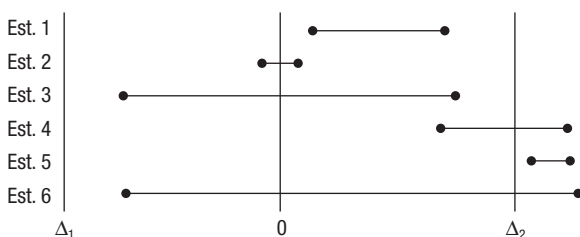


Figura 7-11 Los estudios 1 a 3, que excluyen Δ_1 y Δ_2 , permiten concluir equivalencia.

Ejercicio 7.20



Si en el ejemplo 7.20 de los datos de la figura 7-11 se hubiera hecho el CH para decidir diferencias, ¿en qué estudios de los anteriores se concluiría que los tratamientos son diferentes? Razone posibles discordancias.

Definición

Un tratamiento es *no inferior* a otro si éste no le supera en un cierto valor Δ que hace relevantes las consecuencias.

Recuerde

Los planteamientos de «no inferioridad» y «no superioridad» sólo consideran un límite Δ .

Los planteamientos de equivalencia y de no inferioridad han surgido en las agencias reguladoras del medicamento, que deben decidir si autorizan que un nuevo producto comparta un espacio ocupado previamente por otro producto.

Recuerde

Equivalencia y no inferioridad son conceptos de decisión y, por tanto, de contraste de hipótesis.

Tanto la no inferioridad como la no superioridad se establecen cada una mediante un contraste unilateral. Y la equivalencia puede establecerse mediante el uso simultáneo de ambos, por lo que el procedimiento que se utiliza recibe el nombre de contraste doblemente unilateral (CDU) o *Two-one-sided test*.

Ambos contrastes del CDU suelen realizarse con riesgo $\alpha = 0,05$, y el riesgo α global del CDU se mantiene en $0,05$. Si el IC se calcula con una confianza $1 - 2\alpha$ ($0,90$, si $\alpha = 0,05$), coincidirán las conclusiones del CDU con las del IC.

Nota técnica

Aunque la estimación por intervalo se realiza con una confianza $1 - 2\alpha = 0,90$, el criterio de decisión basado en dicho IC tendrá un riesgo $\alpha = 0,05$. Ello es así porque ambos límites de no equivalencia (que definen las dos H_0) no pueden ser simultáneamente ciertos; y, por tanto, los riesgos α con que se ha contrastado cada una no necesitan ser sumados.

Ejemplo 7.21

Un nuevo (N) antiinflamatorio tiene una tolerabilidad superior a cierto producto clásico de referencia (R). Interesa poder demostrar que sus niveles de eficacia son parecidos para poder compartir el mercado. La eficacia se mide por la proporción de casos en los que desaparece el dolor a los 30 min. Ambos fármacos serán equivalentes en eficacia si las proporciones de desaparición del dolor no difieren en más de un 8%. El intervalo de confianza (90%) de la dife-

Ejemplo 7.21 (Cont.)

rencia de ambas proporciones oscila entre -6 y $+3\%$. Dado que no alcanza los límites de la no equivalencia, se puede rechazar esta última ($\alpha = 0,05$) y autorizar (A_1) la comercialización de N.

Ejercicio 7.21

Mediante un diseño en que todos los casos pasan por ambos tratamientos, se ha obtenido en $n = 20$ casos el valor de presión arterial diastólica (PAD) tras 3 meses con el tratamiento de referencia y con el nuevo. Se ha obtenido la variable diferencia entre ambas PAD. Se ha establecido el límite de no equivalencia clínica de un hipotensor en ± 10 mmHg. Los resultados han sido $\bar{X}_D = 3$ y $S_D = 10$. Calcule el intervalo de confianza y decida si existe equivalencia.

Ejercicio 7.22

Decidir en el ejercicio anterior si existe equivalencia mediante el doble contraste de hipótesis unilateral.

Ejercicio 7.23

(El mismo hipotensor, pero menos casos). Repita los ejercicios 21 y 22, mediante IC y CH, asumiendo que los resultados han sido: $n = 5$; $\bar{X}_D = 3$ y $S_D = 10$.

Ejemplo 7.22

(Muy técnico) Para el establecimiento de equivalencia en biodisponibilidad (o bioequivalencia) se suele requerir que el cociente de los niveles en sangre entre R y N se encuentre entre 0,8 y 1,25; es decir, que ni R puede estar al 80% de N ($80\% = 4/5$), ni N puede estar al 125% de R ($125\% = 5/4$). En concreto, se pide que la media geométrica de dichos cocientes esté entre ambos valores o, lo que es lo mismo, que la media aritmética de la diferencia entre ambos logaritmos se sitúe entre $\ln(0,8) = -0,223$ y $\ln(1,25) = 0,223$. Así, se trabajará con la «diferencia de los logaritmos naturales» (DL), que se corresponde con el logaritmo de los cocientes que se desea mantener, en promedio, entre los dos valores requeridos. En un diseño de datos apareados los resultados han sido: $n = 12$, $\bar{X}_{DL} = 0,1$ y $S_{DL} = 0,2$.

Mediante IC_{90%}:

$$\begin{aligned} \mu_{LR-LN} &\in \bar{X}_{LR-LN} \pm t_{1-2\alpha} S_{LR-LN}/\sqrt{n} \\ \mu_{LR-LN} &\in 0,1 \pm t_{11,0,90} 0,2 / \sqrt{12} \\ \mu_{LR-LN} &\in (-0,004, 0,204) \end{aligned}$$

Ejemplo 7.22 (Cont.)

Mediante $PDU_{\alpha = 0,05}$:

$$\begin{cases} H_{0A} : \mu_{LR-LN} \leq -0,223 \\ H_{1A} : \mu_{LR-LN} > -0,223 \end{cases}$$

$$t_1 = (0,1 - (-0,223)) / (0,2/\sqrt{12}) = 5,595 \rightarrow P < 0,001$$

$$\begin{cases} H_{0B} : \mu_{LR-LN} \geq 0,223 \\ H_{1B} : \mu_{LR-LN} < 0,223 \end{cases}$$

$$t_2 = (0,1 - 0,223) / (0,2/\sqrt{12}) = -2,130 \rightarrow P \approx 0,028$$

Por lo que tanto el IC como la PDU permiten concluir la equivalencia de ambos productos.

Nota técnica

Los planteamientos de equivalencia que se han resuelto en estos ejemplos hacen referencia a la equivalencia en media. Ello implica que un paciente tiene los mismos valores esperados bajo ambos productos en comparación y, por tanto, ambos preparados o productos son igualmente aconsejables para un nuevo paciente (equivalencia poblacional o prescribibilidad). Para que dos preparados se puedan intercambiar en un paciente ya tratado (equivalencia individual o intercambiabilidad) es necesario, además, que no exista interacción entre el preparado y el paciente, es decir: que la diferencia (quizá nula) entre ambos preparados sea la misma para todos los pacientes. La demostración de esta condición ha sido exigida por algunos, lo que ha resultado en una mayor dificultad para la salida al mercado de productos genéricos.

Soluciones a los ejercicios

7.1 La respuesta correcta es la *d*. La *a* es incorrecta, ya que «conocer el valor del parámetro» es el objetivo de la estimación, quizás por intervalo de confianza, no del contraste de hipótesis. La *b* no es correcta, ya que en la prueba de significación la hipótesis forma parte del enunciado del problema y debe siempre ser previa a la obtención de los datos. La *c* no es correcta, ya que se buscan pruebas en contra de la hipótesis *H* que se desea rechazar.

7.2 La respuesta correcta es la *a* ya que debe situarse en *H* aquello que se desea rechazar para así demostrar su complementario.

7.3 El proceso formal de decisión es el siguiente:

- i) Variable: preferencia por A o por B
- ii) Estadístico: proporción *p* que prefieren A
- iii) Hipótesis *H*: $\pi = 0,5$ (ambos fármacos tienen igual preferencia)
- iv) Si *H* es cierta: $p \rightarrow N(\pi, \pi(1 - \pi)/n) = N(0,5, 0,025^2)$.

Premisas: muestra grande $\pi \cdot n > 5$ y $(1 - \pi) \cdot n > 5$

v) $\hat{z} = (p - \pi) / \sqrt{\pi(1 - \pi)/n} = (0,58 - 0,50) / \sqrt{(0,5 \cdot 0,5/400)} = 0,08 / 0,025 = 3,2$
vi) El valor 0,0007 correspondiente a $\hat{z} = 3,2$ se encuentra en la tabla 4-2 en la fila que empieza por 3,2_ y la columna encabezada por __,0. Si se le suma su simétrico (para $Z < -3,2$), se obtiene $P = 0,0014$. Por ello, puede rechazarse, con $P = 0,0014$ que ambos tratamientos sean iguales: A y B no tienen la misma preferencia. [Nota, en la tabla 4-1, en cambio, el valor máximo de tablas es 2,58 que deja por fuera 0,01. Con esta tabla, el valor reportado sería $P < 0,01$]

vii) $IC_{95\%}: \pi \pm Z_{\alpha/2} \sigma_p = 0,58 \pm 1,96 \cdot 0,025 \approx 0,58 \pm 0,05 = [0,53, 0,63]$

La «auténtica» preferencia π por A se encuentra entre 53 y 73%.

7.4 v) $\hat{z} = (p - \pi) / \sqrt{\pi(1 - \pi)/n} = (0,53 - 0,50) / \sqrt{(0,5 \cdot 0,5/400)} = 0,03 / 0,025 = 1,2$
vi) El valor 0,1151 correspondiente a $\hat{z} = 1,2$ se encuentra en la tabla 4-2 en la fila que empieza por 1,2_ y la columna encabezada por __,0. Si se le suma su simétrico (para $Z < -1,2$), se obtiene $P = 0,2302$. Por ello, dado $P = 0,2302 < 0,05$, nada se opone a aceptar que ambos tratamientos tienen la misma preferencia.

vii) $IC_{95\%}: p \pm Z_{\alpha/2} \sigma_p = 0,53 \pm 1,96 \cdot 0,025 \approx 0,53 \pm 0,05 = [0,48, 0,58]$

7.5 Las tres primeras son correctas, la cuarta es una tontería que no tiene nada que ver y la quinta es un error muy habitual de interpretación del valor de *P*, que cuantifica la probabilidad de unos resultados condicionando a una cierta hipótesis, no la probabilidad de que sea cierta una hipótesis condicionando a unos resultados. Más adelante se insistirá en esta distinción.

7.6 $\hat{z} = (\bar{y} - \mu_0) / (s/\sqrt{n}) = (505 - 500) / \sqrt{(100^2/36)} = 0,3$

El valor 0,3821 correspondiente a $\hat{z} = 0,3$ se encuentra en la tabla 4-2 en la fila que empieza por 0,3_ y la columna encabezada por __,0. Si se le suma su simétrico (para $Z < -0,3$), se obtiene $P = 0,7642$. Por ello, nada se opone a aceptar que la media es de 500. Es decir, no se ha logrado demostrar que $\mu > 500$

$IC_{95\%}: \bar{y} \pm Z_{\alpha/2} \sigma/\sqrt{n} = 505 \pm 1,96 \cdot 100 / \sqrt{36} \approx 505 \pm 33 = [472, 538]$

Se sabe, con una confianza del 95%, que μ está entre 472 y 538

$$7.7 \hat{t} = (\bar{y} - \mu_0) / (S/\sqrt{n}) = (79 - 70) / \sqrt{(12^2/16)} = 3$$

Debería buscarse en la tabla 6-1 el valor correspondiente a una t de Student con 15 grados de libertad. Se encuentra que el valor 2,977 de 14 grados de libertad deja por fuera 0,01. La probabilidad que deja fuera el valor 3 será $P < 0,01$ y, por tanto, se ha logrado demostrar que $\mu > 70$. [Nota: Excel proporciona el valor exacto $P = 0,00897274$, pero es más legible $P < 0,01$]

7.8 La correcta es la respuesta c .

7.9 Ahora cambian los niveles P de significación, que deben dividirse por 2: y se obtiene 0,3821 y 0,0045, por lo que las conclusiones no cambian. Nótese que este planteamiento unilateral es más adecuado en estos dos ejemplos.

7.10 Es correcta la respuesta e , ya que se trata de una prueba unilateral cuya H incluye el 0 y todos los valores negativos. Dado que el estadístico se sitúa en H , la única conclusión posible en una prueba de significación es «nada se opone a aceptar H ». Nótese que en un contraste de hipótesis, ambos riesgos, α y β , están acotados, lo que permite tomar ambas decisiones (en una PS , sólo podía rechazarse H).

7.11 Nótese que, al pasar de PS a CH , el enunciado debería pedir una decisión.

En 7.3, al ser $\bar{z} = 3,2 > Z_{\alpha/2} = 1,96$, se rechaza H .

En 7.4, al ser $\bar{z} = 1,2 < Z_{\alpha/2} = 1,96$, nada se opone a aceptar H .

En 7.6, al ser $\bar{z} = 0,3 < Z_{\alpha/2} = 1,96$, nada se opone a aceptar H .

En 7.7, al ser $\bar{z} = 3,0 > t_{14, \alpha/2} = 2,145$, se rechaza H .

7.12 Es correcta la respuesta b .

7.13 Es correcta la respuesta b .

7.14 Efectivamente, las correctas son las tres largas d , h , e l : ¡es peligroso abreviar!

7.15 No, los resultados de su experimento lo único que le dicen es que, asumiendo que los dos productos sean iguales, la probabilidad de obtener unos resultados como los suyos (o más extremos) no es muy pequeña. Por tanto, no puede demostrar que H sea falsa, lo que no equivale a haber demostrado que H sea cierta. Por ello, no puede afirmar que tengan la misma eficacia. Más adelante se estudia cómo puede demostrarse la equivalencia entre dos productos.

7.16 Todas son ciertas.

7.17 La correcta es la respuesta a , si bien formalmente la frase más correcta sería: «por el momento, se acepta tentativamente H , ya que no existen evidencias en su contra».

7.18 Son correctas la a , y la c .

7.19 a y c son claramente inferencia; así como d y e , decisión. b hace inferencia sobre las consecuencias de una decisión.

7.20 Se rechazaría la H de estricta igualdad en los estudios 1, 4 y 5. Nótese que el estudio 1 tiene un IC razonablemente estrecho, que le permite concluir tanto equivalencia (porque excluye Δ_1 y Δ_2) como diferencias (porque excluye 0); es decir, los dos tratamientos no son estrictamente iguales, pero sus diferencias no alcanzan el criterio de relevancia. El estudio 6, en cambio, tiene un IC tan amplio, aporta tan poca información, que no le permite ni rechazar la estricta igualdad ni el límite de relevancia clínica. Los restantes estudios no presentan estas paradojas: el 2 y el 3 no consiguen rechazar la estricta igualdad y sí logran establecer equivalencia (aunque el 2 tiene un IC más estrecho que implica que se dispone de mucha información); y el 4 y el 5 consiguen rechazar la estricta igualdad y no logran establecer equivalencia. Nótese que el estudio 5 no incluye el margen de equivalencia Δ_2 , pero que se sitúa al lado de la no equivalencia (lo que coincide con el planteamiento unilateral).

7.21 $IC_{90\%}: \mu_D \in \bar{x}_D \pm t_{1-2\alpha} \cdot S_D/\sqrt{n}$

$$\mu_D \in 3 \pm t_{19, 0,90} \cdot 10/\sqrt{20} \rightarrow \mu_D \in 3 \pm 1,729 \cdot 10/\sqrt{20} \rightarrow \mu_D \in (-0,866, 6,866).$$

Luego la media de las diferencias entre las presiones de ambos hipotensores está entre $-0,866$ (el de referencia consigue presiones mas bajas en media: gana por $0,866$ mmHg) y $+6,866$ (el nuevo consigue presiones más bajas: gana por $6,866$ mmHg). Luego la diferencia entre ambos fármacos está entre los límites -10 y $+10$: se ha establecido equivalencia.

7.22 CH: $PDU_{\alpha=0,05} \begin{cases} H_{0A}: \mu_D \leq -10 \\ H_{1A}: \mu_D > -10; & t_1 = (3 - (-10))/(10/\sqrt{20}) = 5,814 > 1,645 = z_{0,05} \\ H_{0B}: \mu_D \geq 10 \\ H_{1B}: \mu_D < 10; & t_2 = (3 - 10)/(10/\sqrt{20}) = -3,130 < -1,645 = z_{0,95} \end{cases}$

Con riesgo $\alpha = 0,05$, se aconseja actuar autorizando el genérico.

7.23 $IC_{90\%}: \mu_D \in \bar{X}_D \pm t_{1-2\alpha} \cdot S/\sqrt{n}$

$$\mu_D \in 3 \pm t_{4, 0,90} \cdot 10/\sqrt{5} \rightarrow \mu_D \in 3 \pm 2,132 \cdot 10/\sqrt{5} \rightarrow \mu_D \in (-6,534, 12,534).$$

Ahora, el $IC_{90\%}$ sobrepasa el dintel superior que marca la no equivalencia y, por tanto, no se puede defender que haya equivalencia.

CH: $PDU_{\alpha=0,05} \begin{cases} H_{0A}: \mu_D \leq -10 \\ H_{1A}: \mu_D > -10 & t_1 = (3 - (-10))/(10/\sqrt{5}) = 2,907 \rightarrow P \approx 0,022 < 0,05 = \alpha \\ H_{0B}: \mu_D \geq 10 \\ H_{1B}: \mu_D < 10 & t_2 = (3 - 10)/(10/\sqrt{5}) = -1,565 \rightarrow P \approx 0,096 > 0,05 = \alpha \end{cases}$

Asimismo, aunque la primera prueba aún permite afirmar que la media de las diferencias está por encima de -10 , la segunda no ha permitido establecer que esté por debajo de $+10$. Por tanto, no se ha podido demostrar que la media de las diferencias de ambos hipotensores esté entre -10 y $+10$. Por tanto, la acción debe ser no autorizar el genérico (A_0).