

Capítulo 8:

Comparación de dos grupos

Presentación

Este capítulo expone la popular prueba t de Student para comparar 2 medias, de la que se ofrece una interpretación intuitiva. Se explica por qué el proceso de inferencia estadística requiere un grupo control y las ventajas de concentrar el estudio en el análisis de la media. Finalmente, mediante el método de aparear los datos, se insiste en el concepto de eficiencia estadística.

Objetivos

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

- Percibirá la necesidad de disponer de un grupo control.
- Percibirá la necesidad de que los grupos en comparación difieran sólo en el tratamiento.
- Interpretará la diferencia entre medias como la magnitud del efecto en cada caso.
- Interpretará la amplitud de un intervalo de confianza como la ignorancia sobre el valor del parámetro, no como la variabilidad entre las unidades.
- Interpretará la «razón t » como un cociente «señal/ruido».
- A partir de las medias y de las varianzas de cada grupo, sabrá calcular la prueba t de Student de comparación de dos medias con datos independientes.
- Percibirá las ventajas de que se cumplan las premisas necesarias para realizar la prueba t .
- A partir de la media y de la varianza de la diferencia, sabrá calcular la prueba t de Student de comparación de dos medias con datos apareados.
- Sabrá que el apareamiento permite aumentar la precisión de la estimación del efecto del tratamiento.
- Decidirá que dos muestras están apareadas si los valores de las parejas son similares.

La importancia de tener un control

En los capítulos anteriores se ha visto cómo inferir los resultados desde una muestra a una población. Se ha aprendido a realizar intervalos de confianza y pruebas de hipótesis sobre la media o la probabilidad. Pero sólo se ha expuesto el estudio de una muestra, lo que limita las posibilidades de comparar, por ejemplo, resultados entre dos procedimientos. Incluso en la evaluación de un solo procedimiento, conviene tener un grupo control que proporcione una referencia a la comparación.

Ejemplo 8.1



Un profesor, preocupado por la gran «bolsa» de repetidores que tiene su asignatura, introduce una novedad pedagógica en su docencia: un programa informático de autoaprendizaje que permite al alumno ir comprobando su nivel de progreso en la resolución de problemas. Para medir el efecto de dicho programa, evalúa a los repetidores al inicio y al final del curso y estima, definiendo como variable respuesta Y , la diferencia de conocimientos entre el final y el inicio. El $IC_{95\%}$ de la media de este incremento es de entre 2 y 3 puntos, por lo que concluye que su sistema tiene esa eficacia.

Falso. O, por lo menos, no se sostiene, ya que existen muchas otras variables que pueden ser explicación alternativa a estas diferencias. Especialmente, todo lo que los repetidores hayan estudiado por otros métodos. O incluso, ¿cómo garantizar que el segundo examen no era más fácil que el primero?

Por otro lado, la pregunta de interés práctico es, en general, la elección entre dos opciones alternativas. Puede tener interés, a nivel teórico, conocer el rendimiento o el coste de un sistema; pero, si no hay alternativa, se utilizará el único sistema disponible, por lo que la pregunta no tiene interés práctico. Al menos, la opción alternativa debería ser no hacer nada: o bien implementar el sistema nuevo o bien dejarlo todo igual. Una vez se han definido las dos posibilidades que se desean comparar, un buen diseño de recogida de datos debe informar sobre ambas.

Recuerde



La pregunta más interesante incluye la comparación con una alternativa de control.

Ejemplo 8.1 (Cont.)

El profesor anterior ha seleccionado a los repetidores que cumplían ciertos criterios de homogeneidad y en los que creía que dicho sistema aún podía tener efecto porque no eran situaciones especiales (p. ej., era la primera vez que repetían). A estos estudiantes les solicita permiso para participar en una experiencia docente y requiere su compromiso para seguir las normas de la misma. Los que aceptan son repartidos al azar en dos grupos, uno con acceso al sistema nuevo y el otro, sin. Ambos grupos pasan, en idénticas condiciones, la evaluación inicial y la final, que son corregidas de forma automática. Define como variable de interés el incremento de conocimiento, definido como la diferencia entre la evaluación final y la inicial. El profesor calcula el intervalo de confianza de la diferencia entre los que tienen acceso al sistema respecto a los que no lo tienen. Igual que antes (pero comparando con el otro grupo), el $IC_{95\%}$ de $\mu_1 - \mu_2$ muestra en el grupo experimental un incremento que supera al del grupo control entre 2 y 3 puntos. Este resultado es mucho más sólido. Se cree, con una confianza del 95%, que utilizar este programa mejorará entre 2 y 3 puntos el rendimiento de los repetidores.

Comentario

«Más sólido» pero no definitivo, porque (aunque menos que antes) aún se podría encontrar algún punto delicado. Por ejemplo, si el efecto ha sido realmente debido al sistema de autoaprendizaje o puede ser explicado por la motivación extra de sentirse «agraciado» en el reparto de los grupos. En este último caso, la mejora observada no se reproduciría en el futuro, al incorporar el sistema al aprendizaje habitual.

Recuerde

Al comparar dos tratamientos, conviene que el tratamiento recibido sea la única diferencia entre los dos grupos.

Ejemplo 8.2

Un servicio de traumatología decide evaluar el efecto de implantar una nueva prótesis sustitutiva tras la intervención. Para valorarlo, comparan la media de la calidad de vida de unos pacientes del año anterior (previa a la nueva prótesis) con la de otros pacientes del año siguiente. Los resultados son estadísticamente significativos y favorecen

Ejemplo 8.2 (Cont.)

al nuevo planteamiento. ¿Han demostrado que el nuevo sistema es mejor? No, aunque sean resultados alentadores no son, ni mucho menos, definitivos. Tres grupos de críticas serían las siguientes:

- 1) ¿Son comparables las características de los pacientes de los dos años?
 - 2) ¿Han habido otros cambios de un año a otro? ¿Los profesionales han mejorado sus habilidades? ¿Se han introducido también otras terapias?
 - 3) ¿Qué significa esta calidad de vida? ¿Puede haber sido valorada de forma más generosa en el segundo año?
- Nótese que estos tres grupos de preguntas están buscando terceras variables que puedan ser una explicación alternativa de las diferencias encontradas y, por lo tanto, estén sesgando la comparación en el sentido de generar grupos que difieren en algo más que la intervención en estudio.

Recuerde

Al comparar dos tratamientos, las tres preguntas clave son:

- 1) *¿Son iguales los grupos al inicio?*
- 2) *¿Se ha podido introducir alguna diferencia durante el período experimental?*
- 3) *La evaluación, ¿ha podido introducir otras diferencias?*

Ejercicio 8.1

¿Cuáles son las respuestas ideales (SÍ/NO) a estos tres grupos de preguntas?

Ejercicio 8.2

Proponga un diseño que permita evaluar los efectos de la novedad terapéutica del ejemplo 8.2.

Recuerde

Dos intervenciones que se deseen comparar deben tener:

- 1) *Dos grupos iniciales iguales.*
- 2) *El seguimiento idéntico.*
- 3) *La misma evaluación.*

La importancia de estudiar la medias

Si se dispone de una variable con distribución normal, el análisis habitual es la comparación de las medias en los dos grupos de tratamiento. Pero ¿por qué se comparan las medias si luego se aplicará el nuevo tratamiento a las unidades? ¿Qué información aportan estas medias sobre lo que pasa en cada unidad? Veamos las razones.

Supóngase que el procedimiento nuevo que se desea comparar con el estándar tiene un efecto Δ que es el mismo en todas las unidades. El resultado de aplicar el nuevo procedimiento provocará una traslación de la distribución original exactamente igual a Δ , como muestra la figura 8-1.

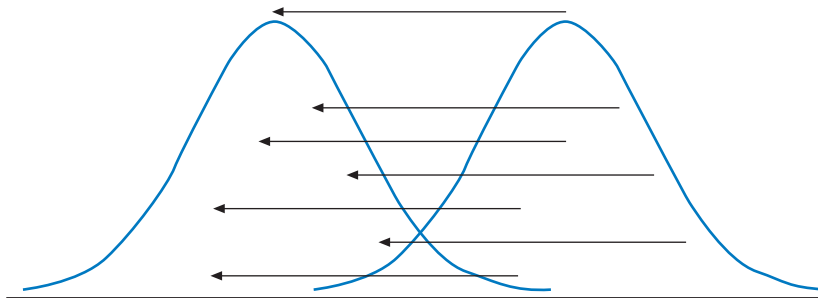


Figura 8-1 Comparación de dos distribuciones normales. Efecto constante, aditivo.

Ejemplo 8.3



Cierto fármaco antihipertensivo desciende la presión sistólica en exactamente 10 mmHg. Este descenso es el mismo en todos los pacientes, sea cual sea su nivel anterior de presión. Los pacientes siguen teniendo diferentes valores de presión sistólica, pero porque ya eran diferentes antes del tratamiento, no porque ésta haya incluido variabilidad en el proceso. Nótese que, sea cual sea la forma de la distribución de la presión sistólica antes del tratamiento, esta distribución se trasladará exactamente 10 unidades, pero conservando la misma forma y la misma dispersión.

Recuerde



Si el efecto Δ es el mismo en todas las unidades, basta con estimar la diferencia entre las medias de ambas poblaciones para conocer el efecto que se dará en cada unidad de esa población.

Nota técnica

Puede demostrarse que la manera más **eficiente** (más precisa) de comparar la posición de dos variables con distribución normal es, precisamente, mediante la comparación de sus medias. De aquí la importancia de comparar las medias: si se asume que la forma de ambas distribuciones es idéntica, es el procedimiento estadístico más eficiente para estimar un efecto único que se observará en cada una de las unidades.

Una consecuencia muy importante de este efecto común en todos los casos es que la distribución es la misma bajo el procedimiento estándar y bajo el nuevo. Por lo tanto, también será igual su varianza.

Nota técnica

Esta situación de igual varianza recibe el nombre técnico de **homocedasticidad**. Permite creer que la diferencia entre las medias está estimando el efecto en cada unidad de la población.

Ejercicio 8.3

Al comparar el efecto de dos tratamientos A y B, se obtiene un intervalo del 95% de confianza de las diferencias de sus medias que va de 5 a 8. Se cree, con una confianza del 95% que:

- La auténtica diferencia de medias poblacionales se encuentra entre 5 y 8.
- El efecto diferencial de A respecto a B consiste en descender entre 5 y 8 unidades más.
- En el 95% de las medias muestrales la diferencia se encuentra entre 5 y 8.
- En el 95% de los pacientes, la diferencia del efecto está entre 5 y 8.

Ejercicio 8.4

En un estudio comparativo entre dos terapias alternativas A y B, el $IC_{95\%}$ de la diferencia de las medias de los dos grupos dice que el descenso de la carga viral ha sido entre 10 y 100 copias superior en el grupo tratado con A. Esto significa que (elija una):

- En los casos en los que el tratamiento A ha sido más favorable, en el grupo A se ha conseguido un descenso de 100 unidades más que en el grupo B, mientras que en los menos favorables ha sido de 10 copias.
- En los respondedores, la ventaja de A ha sido de 100 copias, mientras que en los no respondedores, ha sido de 10 copias.

Ejercicio 8.4 (Cont.)

- c) A descende más que B un valor único y común para todos los casos, pero que se desconoce y debe ser uno de los comprendidos entre 10 y 100 copias.
d) Ninguna de las anteriores es correcta.

Recuerde

La amplitud del intervalo de confianza refleja el grado de ignorancia, no la variabilidad del efecto.

Nota técnica

En el entorno de comparación de medias no es correcto hablar de «respondedores» y «no respondedores». Si se sospecha que puede haber dos grandes grupos de pacientes, los que responden y los que no (respuesta dicotómica), un análisis correcto podría ser la comparación de la proporción de respondedores entre los grupos.

Resumen

Si la distribución en los dos grupos tiene la misma forma y la misma dispersión, la comparación de dos poblaciones se reduce a la comparación de sus medias.

Contraejemplo 8.4

Supóngase ahora que el efecto del tratamiento consiste en reducir un 20% la presión sistólica, de forma que a un paciente con 100 mmHg se la baja a 80 mmHg y a uno de 150 mmHg se la baja a 120 mmHg. Ahora, este efecto «proporcional» no respeta la variabilidad de las observaciones, ya que provoca mayor descenso cuanto mayor es la presión inicial, como puede verse en la figura 8-2.

Ahora, la diferencia entre medias ya no representa el efecto en cada unidad, lo que podría complicar mucho la interpretación: como el efecto dependería del valor inicial debería estimarse por separado según la severidad de la presión inicial. Afortunadamente, las matemáticas acuden en nuestra ayuda: una simple transformación logarítmica (la misma que define el pH) permitirá convertir este modelo multiplicativo («descenso proporcional») en un modelo aditivo en el que poder utilizar la diferencia de las medias.

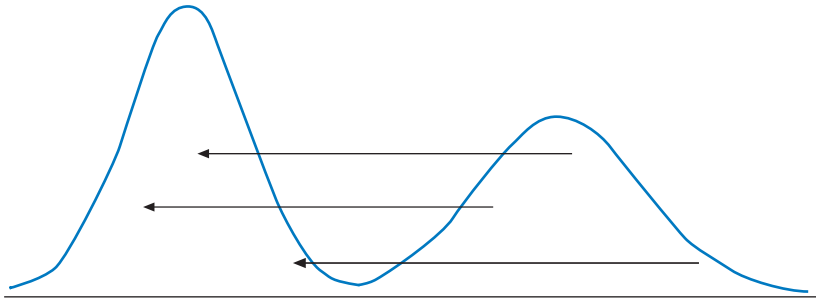


Figura 8-2 Comparación de dos distribuciones normales. Efecto no constante.

Comparación de dos medias en muestras independientes. Varianzas iguales

Veamos ahora cuál es la solución numérica de la comparación de dos medias. Para empezar, se asume que se conocen las varianzas y que las distribuciones de las dos variables en comparación son normales. Se asume también que los procesos de muestreo de los tratamientos en comparación son aleatorios e independientes.

Para estudiar la **diferencia entre las medias poblacionales** μ_1 y μ_2 se recurre al estadístico $(\bar{y}_1 - \bar{y}_2)$ o **diferencia de las medias muestrales**.

¿Qué **distribución** sigue este estadístico?: **normal**, ya que es una combinación de variables normales. Sabiendo la forma, queda por especificar su centro y su dispersión: ¿qué valen su esperanza y su varianza? La **esperanza** de la diferencia de las medias muestrales es directamente la diferencia de sus esperanzas o medias poblacionales:

$$E(\bar{y}_1 - \bar{y}_2) = E(\bar{y}_1) - E(\bar{y}_2) = \mu_1 - \mu_2$$

La diferencia entre las medias muestrales $(\bar{y}_1 - \bar{y}_2)$ estima, sin sesgo, cuánto vale la diferencia de las medias poblacionales $(\mu_1 - \mu_2)$.

La **varianza** de la diferencia de las medias muestrales se convierte, en esta situación de muestras independientes, en la suma de las varianzas de las medias:

$$V(\bar{y}_1 - \bar{y}_2) = V(\bar{y}_1) + V(\bar{y}_2) = \sigma_1^2/n_1 + \sigma_2^2/n_2$$

Comentario



La varianza de una resta es la suma de las varianzas. Se puede imaginar este resultado si se tiene en cuenta que se trata de variables aleatorias independientes y, para conocer su diferencia, las oscilaciones aleatorias de cada variable contribuyen al ruido de estimación. Por ejemplo, en un primer experimento, la media muestral del grupo 1 puede oscilar hacia arriba, y la del grupo 2, independiente del anterior, hacia abajo. En un segundo experimento, puede ser al revés o similar... Ambos errores aleatorios se añaden en la estimación de la diferencia de las medias poblacionales $\mu_1 - \mu_2$.

Recuerde



La diferencia de las medias muestrales sigue una distribución normal cuyo centro es la diferencia de las medias poblacionales y cuya dispersión es la suma de la oscilación de cada media.

$$(\bar{y}_1 - \bar{y}_2) \rightarrow N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$$

Una vez conocida la distribución de la diferencia de las medias muestrales, ya se puede proceder de forma similar a como se hacía para el caso de una muestra.

Si se asume que se conocen las varianzas poblacionales, se puede tipificar esta diferencia de medias en forma de cociente señal/ruido:

$$\hat{z} = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

Y se puede realizar inferencia, bien mediante un intervalo de confianza:

$$IC_{1-\alpha}(\mu_1 - \mu_2): (\mu_1 - \mu_2) \in (\bar{y}_1 - \bar{y}_2) \pm Z_{\alpha/2} \cdot \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$$

bien mediante un contraste de hipótesis:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 - \mu_2 = \Delta \end{cases}$$

Bajo H_0 :

$$\hat{z} = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

Como usualmente σ_1 y σ_2 son desconocidas, deberán ser estimadas. Si se puede asumir igualdad de varianzas ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) entonces, S_1^2 y S_2^2 serán estimadores del mismo parámetro σ^2 , por lo que pueden ser combinados en un **estimador único ponderado** por la información que aporta cada uno:

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2}$$

Nota técnica



Si ambas muestras son del mismo tamaño, este estimador S^2 se reduce al promedio de las varianzas de cada muestra:

$$S^2 = \frac{(S_1^2 + S_2^2)}{2}$$

Este estimador promedia el cuadrado de todas las desviaciones con la media, dándoles el mismo peso vengan del grupo que vengan. Para cada grupo, utiliza su media.

Finalmente, se llega al conocido estadístico llamado «razón t o t ratio»:

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow N(0,1)$$

En el numerador está la diferencia entre las medias muestrales, o *señal* proporcionada por los datos. En el denominador está la raíz de la varianza del estimador que figura en el numerador, es decir, el error típico de esta señal: su *ruido* aleatorio.

Recuerde



La razón t es el cociente entre la señal y el ruido.

El estadístico *razón t* sigue una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad. Las premisas necesarias son: 1) variable, con distribución normal; 2) varianzas iguales, y 3) muestras independientes.

Nota técnica



Aunque la variable original en estudio no siga la distribución normal, la aproximación se considera razonable si el tamaño de ambas muestras supera los 30 casos. Cuanto más se asemeje a la distribución normal la variable en estudio más pequeño es este número requerido de casos.

Ejercicio 8.5



- El estadístico «razón t » sigue una t de Student con
- $n_1 + n_2 - 2$ grados de libertad.
 - $n - 1$ grados de libertad.
 - 30 grados de libertad.
 - Este estadístico no sigue una t de Student.

Ejercicio 8.6



- Para que el estadístico «razón t » siga una t de Student con $n_1 + n_2 - 2$ grados de libertad, las premisas necesarias son:
- MAS e independientes.
 - Normalidad de la variable original.
 - Homocedasticidad o igualdad de varianzas.
 - Todas son correctas.

Ejemplo 8.5

Para tomar la decisión de sustituir el tratamiento 2 por el tratamiento 1, se ha realizado un ensayo clínico ($n_1 = 50$ y $n_2 = 100$) cuyos resultados han sido: $\bar{y}_1 = 24$ e $\bar{y}_2 = 21$, siendo $s_1 = 8$ y $s_2 = 6$. Presuponiendo que las varianzas poblacionales son iguales, ¿se puede considerar que ambos tratamientos inducen la misma calidad de vida?

Solución:

- 1) $\begin{cases} H_0 : \mu_1 = \mu_2 \text{ (misma calidad de vida)} \\ H_1 : \mu_1 - \mu_2 = \Delta \text{ (El tratamiento 1 es superior en } \Delta \text{ unidades)} \end{cases}$

- 2) **Estadístico:**

$$\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- 3) **Distribución** bajo $H_0 : \hat{t} \rightarrow t_{n_1 + n_2 - 2}$

Premisas: Y_1, Y_2 normales;

$\sigma_1^2 = \sigma_2^2$ MAS e independientes

- 4) **Regla decisión:** rechazar H_0 ($\alpha = 0,05$)

si $\hat{t} > t_{n_0 + n_0 - 2, 0,975} = 1,976$

- 5) **Cálculo:** $S^2 = (8^2 \cdot 49 + 6^2 \cdot 99) / 148 \approx 44$

$$\hat{t} = (24 - 21) / \sqrt{[(44/50) + (44/100)]} \approx 2,61$$

La señal es más del doble del ruido aleatorio.

$$P(|t_{n_1+n_2-2}| > 12,61) < 0,01$$

- 6) **Decisión:** dado que $2,61 > 1,976$

se rechaza que $\mu_1 = \mu_2$ con riesgo $\alpha = 0,05$

- 7) $IC_{95\%}(\mu_1 - \mu_2) = (\bar{y}_1 - \bar{y}_2) \pm 1,976 \sqrt{(S^2/n_1 + S^2/n_2)}$

$$= 3 \pm 1,976 \cdot 1,149 = 3 \pm 2,27 = [0,73, 5,27]$$

- 8) **Conclusión** práctica: el tratamiento 1 tiene una calidad de vida entre 0,73 y 5,27 puntos mayor.

Sustituiremos el tratamiento 2 por el tratamiento 1.

Recuerde

La «razón t » o cociente señal/ruido empieza a ser relevante a partir de 2 (la señal debe doblar al ruido).

Ejercicio 8.7

Llano et al. (45) comparan dos momentos evolutivos diferentes (SI y NSI) en grupos de pacientes ($n_1 = 56$, $n_2 = 75$) de sida. Las medias de CD4 han sido 125 y 329 y las desviaciones típicas, 155 y 223. Calcule el estimador de la varianza común y estime por intervalo la diferencia media entre ambos tratamientos.

Nota técnica



El estadístico «razón t» es muy importante. Considerémoslo desde otro punto de vista.

$$\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2)}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Ahora es el producto de dos cantidades. La primera representa, como antes, la señal. Pero ahora, al dividirla por la desviación típica ha perdido sus unidades: está estandarizada o tipificada. La segunda cantidad representa la «magnitud» de la muestra.

Así, la razón «t» es la **señal tipificada** multiplicada por la **cantidad de información** que aporta el diseño muestral.

Nota técnica



Para un tamaño muestral total N prefijado ($N = n_1 + n_2$), el máximo de esta «magnitud» es para tamaños muestrales iguales ($n_1 = n_2$). Es decir, si se dispone de n unidades, el diseño de dos muestras independientes más eficiente consiste en hacerlas de igual tamaño.

Recuerde



Si dispone de un total de N unidades, el diseño óptimo, que conduce al IC más estrecho, dispone N/2 en cada grupo.

Comparación de dos medias en muestras independientes. Varianzas diferentes

Si no se puede asumir que las varianzas sean iguales aparecen dificultades.

Comentario



La primera dificultad es que la diferencia entre las medias ya no puede representar al efecto «común» en cada uno de los casos. Lo único que puede representar es cierto efecto promedio y pierde utilidad, ya que aparece incertidumbre sobre cuál será el efecto en una unidad.

La segunda dificultad es más teórica: resulta que el estadístico, análogo al anterior, que ahora utilizaríamos:

$$\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

ya no sigue una t de Student, pues el denominador no es ahora una sola variable, sino dos.

La mejor solución consiste en encontrar una transformación de Y, en la que las varianzas sean iguales (y la distribución, normal), ya que se resuelven, a la vez, varios problemas.

Nota técnica



Podría darse la situación en la que el efecto en estudio no fuera aditivo (una constante que se suma o resta a cada observación) sino proporcional, multiplicativo como en el contraejemplo 8.4 del tratamiento que reducía la presión en un 20%. Ahora, la constante que permite comparar ambas distribuciones debería obtenerse mediante un cociente en lugar de una resta. Vimos que tomar logaritmos permite convertir ese cociente en una resta. Además, debido a que la normal aparece cuando la variable en estudio es la suma, no el producto, de pequeños efectos, al aplicar logaritmos, la distribución resultante suele aproximarse mejor a una normal.

Nota técnica



Existen muchas soluciones al problema de cuál es la distribución del estadístico señal/ruido en caso de heterocedasticidad. Permiten calcular el nivel P de significación y el intervalo de confianza, pero recuerde que la interpretación ya no será tan directa, porque el valor estimado en el IC ya no puede representar el efecto en cada paciente, sino a un cierto efecto «promedio».

Comparación de dos medias con muestras dependientes o apareadas

Supóngase que, para comparar 2 tratamientos, A y B, los pacientes en que se comparan son los mismos. Es decir: en cada unidad se obtiene información del rendimiento de ambos tratamientos. Así, se consigue eliminar, en la comparación de los tratamientos, el efecto que podría producir el hecho de comparar diferentes pacientes. Cada unidad proporciona información sobre la diferencia del efecto de ambos tratamientos y puede definirse una nueva variable D, que sea esta diferencia.

$$D_i = Y_{iA} - Y_{iB}$$

Comentario



La comparación se hace «dentro» de cada unidad: de observar diferencias entre las medias de ambos tratamientos, no podrían ser explicadas por diferencias entre los pacientes. En otras palabras, se suprime la posibilidad de que diferencias entre las unidades con las que se evalúa cada tratamiento, sean una explicación alternativa que impida poder atribuir a los tratamientos las diferencias observadas.

Es instructivo descomponer la varianza σ^2 de la variable que mide el efecto en dos componentes: las diferencias entre los distintos pacientes (**variabilidad entre unidades** o efecto de las unidades) y las divergencias entre diferentes determinaciones en un mismo paciente, representando la **variabilidad intra unidad**, que incluye también el error de medida.

En un diseño apareado la comparación se hace dentro de cada unidad, por lo que se elimina la varianza debida a las unidades, lo que hace al diseño más eficiente.

Comentario



Dado que puede esperarse de un proceso de medida razonable que el error que se comete al determinar el auténtico valor sea muy inferior a las diferencias entre las unidades, el diseño con datos apareados elimina la fuente de variación más importante (entre). Como ejemplo, supongamos la determinación de la altura de una persona con un procedimiento muy poco preciso: la cinta métrica «de costurera». Es poco preciso porque determinaciones repetidas pueden proporcionar valores diferentes. Aun así, las diferencias entre personas serán siempre mucho mayores que las variaciones entre los diferentes valores obtenidos en la misma persona por determinaciones repetidas.

El procedimiento estadístico que se utiliza en datos apareados es muy sencillo si se define la variable diferencia. De esta manera, la prueba de la hipótesis de que ambos tratamientos son iguales se reduce a la prueba de conformidad de una media de una muestra, de la que ya se vieron ejemplos en el capítulo anterior.

Ejemplo 8.6



En 6 pacientes se han estudiado dos tratamientos A, B. Se desea decidir si tienen medias iguales. Los resultados han sido:

							\bar{y}_j	S_j^2	S^2
Y_{IA}	23,05	39,06	21,72	24,47	28,56	27,58	27,407	39,421	42,023
Y_{IB}	20,91	37,21	19,29	19,95	25,32	24,07	24,458	44,625	

La solución **errónea**, mediante la prueba de datos independientes, sería:

$$\hat{t} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(27,407 - 24,458)}{\sqrt{42,023 \left(\frac{1}{6} + \frac{1}{6} \right)}} = \frac{2,948}{3,743} \approx 0,788$$

siendo $42,023 \approx (39,421 + 44,625) / 2$

Como $0,788 < 2,228 = t_{10, 0,05}$ nada se opone a aceptar H_0 en esta solución errónea.

La solución **correcta** de la prueba de datos dependientes o apareados empieza calculando la diferencia $D_i = Y_{IA} - Y_{IB}$:

Ejemplo 8.6 (Cont.)

La solución **correcta** de la prueba de datos dependientes o apareados empieza calculando la diferencia $D_i = Y_{iA} - Y_{iB}$:

							\bar{Y}_j	S_j^2	S^2
Y_{iA}	23,05	39,06	21,72	24,47	28,56	27,58	27,407	39,421	42,023
Y_{iB}	20,91	37,21	19,29	19,95	25,32	24,07	24,458	44,625	
$D_i = Y_{iA} - Y_{iB}$	2,14	1,85	2,43	4,52	3,24	3,51	2,948	1	

$$\hat{t} = (\bar{D} - \mu_0) / (s_D / \sqrt{n}) = 2,948 / \sqrt{1/6} = 2,948 / 0,408 = 7,225$$

Como $7,225 > 2,571 = t_{5, 0,005}$ se rechaza H_0 .

A pesar de que hay la misma señal (la media de las diferencias es la diferencia de las medias), ahora el cociente señal/ruido toma un valor muy superior. La varianza de la variable diferencia (1) es muy inferior a la varianza anterior (42,023), que indica el beneficio de hacer un diseño con datos apareados: a pesar de que el numerador (*señal*) es el mismo, el denominador (*ruido*) es muy inferior en la segunda. Ello provoca que la primera «solución», la errónea, no pueda llegar a rechazar la hipótesis nula, pero que la segunda, a pesar de tener la misma señal, 2,948, encuentre pruebas en contra de la hipótesis nula y la pueda rechazar.

Recuerde

Controlar el efecto unidad, estudiando los tratamientos en los mismos casos conlleva dos ventajas metodológicas: a) evita que diferencias entre unidades puedan ser explicación alternativa de las diferencias observadas; y b) aumenta la eficiencia de la estimación al reducir el error aleatorio.

Nota técnica

La varianza de la estimación de la diferencia entre medias incluye un término que resta la covarianza entre ellas. Si se trataba de muestras independientes, esta covarianza se anulaba.

$$V(\bar{y}_1 - \bar{y}_2) = V(\bar{y}_1) + V(\bar{y}_2) - 2\text{Cov}(\bar{y}_1, \bar{y}_2)$$

Esta covarianza representa el efecto individuo, la variabilidad «entre casos». Cuanto más similares son las dos determinaciones y_1 , y_2 en cada individuo, mayor es la ventaja por utilizar un diseño apareado.

Los ejemplos de datos apareados que se han presentado se referían todos al estudio de dos tratamientos en las mismas unidades. Pero pueden plantearse diseños con un menor grado de dependencia o apareamiento. Por ejemplo, y de más a menos: en hermanos gemelos; en hermanos; en dos pacientes de la misma edad, género y antecedentes; en dos pacientes de la misma edad, etc. Ello puede hacer difícil la decisión de si los datos están o no apareados.

Comentario



¿En qué circunstancias se puede creer que un diseño es de datos apareados? La primera pista es el tamaño de la muestra de las dos variables en estudio, ya que para que sean datos apareados se necesita que $n_1 = n_2$. La segunda pista es que exista un nexo que permita calcular las diferencias. Y en tercer lugar, la pregunta clave es si esta conexión entre los pares de observaciones está eliminando alguna variabilidad en el modelo, alguna diferencia natural entre las unidades estudiadas. Dicho de otra manera: ¿cabe esperar que las dos observaciones de la misma pareja sean más similares entre sí que otras dos observaciones seleccionadas al azar? Es decir, conocido el valor de la primera observación, ¿es posible acercarse, adivinar el valor de la segunda? Más formalmente: ¿vale 0 su correlación?

La figura 8-3 muestra la elevada **correlación** entre las variables Y_A e Y_B del ejemplo anterior. En efecto, a mayores valores de la variable Y_A , le corresponden mayores valores de la variable Y_B . El hecho de que todas las observaciones estén por debajo de la diagonal (bisectriz, dado el escalado) muestra gráficamente que los valores de Y_B son cada uno de ellos menores que los de Y_A .

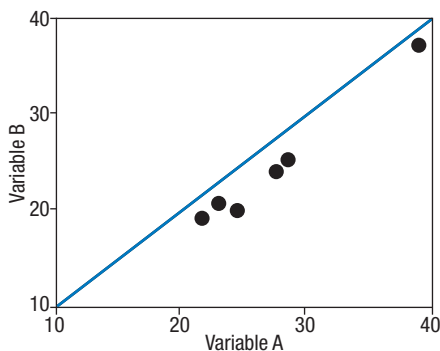


Figura 8-3 En datos apareados es posible estudiar la similitud entre ambas determinaciones.

Nota técnica



La **covarianza poblacional** entre las variables X e Y se define como:

$$\text{Cov}(X, Y) = E[X - E(X)][Y - E(Y)] = E(XY) - E(X)E(Y)$$

Nota técnica



Y la **covarianza muestral** se calcula como

$$\Sigma_{AB} = [\Sigma_{AB} - \Sigma_A \Sigma_B / n] / (n - 1) = [4229,543 - 164,440 \cdot 146,750 / 6] / 5 = 41,523$$

De donde la correlación vale:

$$r_{AB} = S_{AB} / S_A S_B = 41,523 / (6,279 \cdot 6,680) = 0,99$$

Esta **correlación** tan próxima a 1 está indicando que ambas variables van extremadamente juntas: hay un efecto paciente que origina que si el tratamiento A obtiene, por ejemplo, valores más elevados en un paciente, también se obtendrán valores más elevados en ese paciente con el tratamiento B. Puede demostrarse que la covarianza entre ambas variables coincide con la varianza entre unidades.

Ejemplo 8.7



Se desea averiguar si la frecuencia de relaciones sexuales modifica la probabilidad de reinfección por el virus del sida. Se dispone de una base de 120 pacientes de sida de transmisión sexual, de los que 6 se han reinfectado. A cada uno de ellos se le ha buscado en la base de datos un control lo más similar posible de acuerdo con los siguientes criterios: edad, género, hábitos sexuales y características clínicas. Finalmente, se ha escogido aquel control que tuviera un tiempo de seguimiento igual al tiempo de reinfección del caso.

Ejercicio 8.8



En la tabla siguiente figura la frecuencia de relaciones mensuales declaradas por cada caso y cada pareja del ejemplo anterior. Se pide realizar la prueba de datos apareados y la de datos independientes. Concluya en ambas y comente sobre el grado de apareamiento de los casos.

							\bar{y}_i	S_i^2	S^2
Y_{IA}	7,0	4,0	5,0	4,0	11,0	14,0	7,50	17,10	17,48
Y_{IB}	7,0	0,0	9,0	7,0	13,0	8,0	7,33	17,87	
$D_i = Y_{IA} - Y_{IB}$	0,0	4,0	-4,0	-3,0	-2,0	6,0	0,17	16,17	

Nota técnica



Podría darse la situación en la que unas observaciones proporcionen información sobre ambos tratamientos pero otras sólo tengan valores de uno (únicamente cierta porción de los datos está apareada). ¿Por qué se ha perdido esta información? ¿Son comparables los casos que no completan el proceso? Recuerde que es muy peligroso eliminar, sin más, estos casos.

Comentario



Consulte con un estadístico profesional si tiene datos parcialmente apareados.

Ejercicio 8.9



La estimación mediante datos apareados es más eficiente que la de datos independientes porque:

- a) aumenta la diferencia entre las medias;
- b) disminuye la diferencia entre las medias;
- c) aumenta el error típico de estimación;
- d) disminuye el error típico de estimación.

Ejercicio 8.10

La prueba definitiva para saber si unos datos son apareados es:

- a) que los tamaños muestrales sean iguales;
- b) que los datos vengan por parejas;
- c) que las parejas se hayan recogido muy juntas;
- d) que exista cierta similitud o correlación entre los valores de las parejas.

Comparación de dos varianzas con muestras independientes

Comentario



*Se ha visto la comparación de las medias de dos variables con distribución normal. De la misma forma podría verse la comparación de sus varianzas. Pero no lo haremos, especialmente porque creemos que la respuesta fundamental viene de las comparaciones de medias. Se puede argumentar que la igualdad de varianzas es una **premisa** necesaria para la comparación de medias. Y es cierto. Pero la posición mantenida en este texto es que las premisas son previas al establecimiento de la hipótesis y, por tanto, a la obtención de los datos. Si los datos presentan muchas novedades o sorpresas sobre las premisas, da la sensación de que el investigador se ha lanzado a un experimento sin saber muy bien de qué iba. Por otro lado, no se pretende afirmar que las premisas son una verdad absoluta sino razonable: el planteamiento estadístico presentado es **robusto** o sólido ante pequeñas desviaciones de las premisas.*

Soluciones a los ejercicios

8.1 Las respuestas idóneas deben ser SÍ, NO Y NO.

8.2 Cualquier ejemplo similar al del profesor. En los capítulos 11 y 12 se desarrollan con profundidad los aspectos relativos al diseño de experimentos y de observaciones.

8.3 Las respuestas correctas son la *a*) y la *b*), ya que la inferencia estadística hace referencia a los parámetros poblacionales o sus diferencias.

8.4 La respuesta correcta es la *c*). La amplitud del intervalo informa de la magnitud de la ignorancia sobre el efecto concreto, no de una variabilidad de este efecto a lo largo de los casos.

8.5 La respuesta correcta es la *a*).

8.6 La respuesta correcta es la *d*).

8.7 $S^2 = (155^2 \cdot 55 + 223^2 \cdot 74)/129 \approx 38769,9$

$\hat{t} = (125 - 329) / \sqrt{[(38769,9/56) + (38769,9/75)]} \approx -204 / 34,77 \approx -5,87$

Se rechaza la hipótesis de igualdad de medias: el grupo SI tiene valores menores de CD4.

$IC_{95\%} = (125 - 329) \pm t_{129, 0,90} \cdot \sqrt{(S^2/50 + S^2/100)} = -204 \pm 1,657 \cdot 34,1 = -204 \pm 56,5$

$IC_{95\%} = (-260,5, -147,5)$

8.8 Independientes: $S^2 = (17,10 \cdot 5 + 17,87 \cdot 5)/10 \approx 17,48$

$\hat{t} = (7,5 - 7,33) / \sqrt{[(17,48/6) + (17,48/6)]} \approx 0,17 / 2,41 \approx 0,07$

Apareados $\hat{t} = (0,17) / \sqrt{(16,17/6)} \approx 0,17/1,64 \approx 0,10$

Dado que el cociente señal/ruido está muy próximo a cero, no hace falta mirar en tablas.

Ninguna permite rechazar la nula de igualdad: no se han detectado diferencias.

El ruido en la prueba de datos apareados es sólo ligeramente inferior.

Parece que, para esta variable, aparear los datos no aporta grandes beneficios.

8.9 La respuesta correcta es la *d*).

8.10 La respuesta correcta es la *d*).