

14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Clase 7

Pruebas de Hipótesis: Guía Completa

Soluciones

Módulos oficiales del MicroMaster:

M6 (extensión) – Confidence Intervals and Hypothesis Testing, ampliado con χ^2 , F y pruebas de dos muestras

B.Sc. Cristian Lazo Quispe

✉ clazoq@uni.pe

[in linkedin.com/in/cristian-lazo-quispe](https://www.linkedin.com/in/cristian-lazo-quispe) · github.com/CristianLazoQuispe

Material del *Advanced Program in Data Science & Global Skills*, diseñado y desarrollado por **BREIT (Brescia Institute of Technology)**, iniciativa de **Aporta** —plataforma de innovación e impacto social del **Grupo Breca**— en alianza con el **MIT IDSS** (Institute for Data, Systems, and Society).

Basado en MIT 14.310x (E. Duflo & S. Ellison) y en diversas fuentes abiertas (ver bibliografía al final), bajo licencia CC BY-NC-SA 4.0.

11 de agosto de 2026

Niveles de dificultad (*difficulty levels*)

Cada problema está rotulado por color según su nivel, de lo más accesible (Nivel 1) a lo más difícil (Nivel 5). Empieza por donde te sientas cómodo y avanza a tu ritmo; los niveles 4–5 son extensiones para profundizar, no un requisito.

- **Nivel 1 – Fundamentos:** calentamiento muy guiado; para quien parte de casi cero.
- **Nivel 2 – Núcleo:** el nivel objetivo del curso/examen.
- **Nivel 3 – Aplicación:** combina varias ideas en contexto realista.
- **Nivel 4 – Reto:** difícil, integra varios temas.
- **Nivel 5 – Reto+:** muy difícil / fuera del scope típico.

1. Problemas y soluciones

Nivel 1 – Fundamentos (warm-up)

Para quien parte de casi cero: cálculos directos y guiados.

Ejercicio 1.1 (Identificar H_0/H_1 y colas de una frase). Para cada pregunta, escribe H_0 , H_1 y di si la prueba es de una cola o dos:

- (a) ¿La tasa de reprobación *bajó* tras un nuevo método de enseñanza?
- (b) ¿El precio medio de un producto *cambió* respecto al año pasado?
- (c) ¿La velocidad de internet *superó* los 100 Mbps prometidos?
- (d) ¿La proporción de clientes satisfechos *difiere* de 80 %?
- (e) ¿El nuevo fertilizante *aumentó* el rendimiento por hectárea?
- (f) ¿El tiempo de espera en el banco *disminuyó* tras contratar más cajeros?
- (g) ¿Dos métodos de enseñanza producen la *misma* variabilidad en notas?
- (h) ¿La participación en un programa social *es distinta* entre dos regiones?

Solución

(a) $H_0 : p = p_0$, $H_1 : p < p_0$, una cola (izq.). (b) $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$, dos colas. (c) $H_0 : \mu = 100$, $H_1 : \mu > 100$, una cola (der.). (d) $H_0 : p = 0.80$, $H_1 : p \neq 0.80$, dos colas. (e) $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$, una cola (der.). (f) $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$, una cola (izq.). (g) $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 : \sigma_1^2 \neq \sigma_2^2$, dos colas (F -test). (h) $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$, dos colas.

Ejercicio 1.2 (Leer un valor crítico). Con $\alpha = 0.05$: (a) ¿cuál es el valor crítico z para una prueba de dos colas? (b) ¿y para una cola? (c) Con $z_{\text{obs}} = 1.80$, ¿rechazas en una prueba de dos colas? ¿y en una prueba de una cola derecha?

Solución

(a) $z_{0.975} = 1.96$. (b) $z_{0.95} = 1.645$. (c) Dos colas: $1.80 < 1.96 \Rightarrow$ no rechazas. Una cola derecha: $1.80 > 1.645 \Rightarrow$ rechazas. (Mismo dato, distinta conclusión según la dirección de H_1 — por eso hay que fijar la dirección **antes** de ver los datos.)

Ejercicio 1.3 (Calcular un estadístico z). Una fábrica de galletas afirma que el peso medio por paquete es $\mu_0 = 78$ g, con $\sigma = 12$ g (conocida). En $n = 36$ paquetes, $\bar{X} = 82$ g.

- (a) Calcula $z = (\bar{X} - \mu_0)/(\sigma/\sqrt{n})$.
- (b) Con $\alpha = 0.05$ y $H_1 : \mu \neq 78$, ¿rechazas H_0 ?

Solución

(a) $z = \frac{82 - 78}{12/\sqrt{36}} = \frac{4}{2} = 2.0$. (b) $|z| = 2.0 > 1.96 \Rightarrow$ **rechazamos** H_0 .

Ejercicio 1.4 (Leer un p-valor de una figura sombreada). Un estadístico observado da $z_{\text{obs}} = 1.75$.

- (a) Calcula el p-valor de una cola ($\mathbb{P}(Z > 1.75)$).
- (b) Calcula el p-valor de dos colas.
- (c) Con $\alpha = 0.05$, ¿en cuál de los dos casos rechazarías H_0 ?

Solución

(a) p-valor (una cola) $= 1 - \Phi(1.75) \approx 0.0401$. (b) p-valor (dos colas) $= 2(0.0401) \approx 0.0801$. (c) Una cola: $0.0401 < 0.05 \Rightarrow$ rechazas. Dos colas: $0.0801 > 0.05 \Rightarrow$ no rechazas. (El mismo z_{obs} puede llevar a decisiones distintas según la dirección de H_1 .)

Ejercicio 1.5 (¿Pareado o independiente?). Para cada caso, di si el diseño es pareado o independiente:

- (a) Se mide el colesterol de 30 pacientes antes y después de un tratamiento.
- (b) Se compara el ingreso medio de 50 trabajadores de la costa contra 50 de la sierra.
- (c) Se mide el tiempo de reacción de cada persona con la mano izquierda y con la derecha.

Solución

(a) Pareado (mismos pacientes, dos mediciones). (b) Independiente (grupos distintos de personas). (c) Pareado (misma persona, dos mediciones).

Ejercicio 1.6 (Conteo esperado en una tabla de contingencia). En una tabla 2×2 , la fila “Educación superior” tiene un total de 80; la columna “Empleo formal” tiene un total de 90; el total general es 200. Calcula el conteo esperado de la celda (Educación superior, Empleo formal) bajo independencia.

Solución

$$E = \frac{(\text{total fila})(\text{total columna})}{\text{total general}} = \frac{80 \times 90}{200} = 36.$$

Ejercicio 1.7 (Reconocer la familia de prueba (sin resolver)). Para cada escenario, di **solo** qué familia de prueba usarías (no lo resuelvas): (a) comparar la variabilidad de dos máquinas; (b) comparar el ingreso promedio de un grupo tratamiento vs. uno control (RCT, grupos distintos); (c) verificar si la distribución de un dado coincide con un dado justo; (d) probar si un puntaje promedio superó una meta fija, con σ desconocida.

Solución

(a) F -test. (b) t de dos muestras independientes (pooled o Welch, según varianzas). (c) χ^2 de bondad de ajuste. (d) t de una muestra.

Nivel 2 – Núcleo (core)

El nivel objetivo del curso/examen; todos deberían llegar aquí.

Ejercicio 1.8 (Test z completo, dos colas). Una línea de producción llena botellas con una media histórica de $\mu_0 = 750$ ml, $\sigma = 8$ ml (conocida). En $n = 64$ botellas, $\bar{X} = 746.5$ ml. ¿Cambió el llenado medio? ($\alpha = 0.05$.)

Solución

$H_0 : \mu = 750$ vs. $H_1 : \mu \neq 750$. $z = \frac{746.5 - 750}{8/\sqrt{64}} = \frac{-3.5}{1} = -3.5$. Crítico ± 1.96 . Como $|-3.5| = 3.5 > 1.96$, **rechazamos** H_0 ; p-valor ≈ 0.0005 : el llenado medio sí cambió.

Ejercicio 1.9 (Test t completo, una cola). Un gimnasio afirma que sus clientes pierden en promedio $\mu_0 = 25$ h de sueño recuperadas... (contexto: en realidad, mide horas semanales de ejercicio ganadas tras un programa, comparado con la meta $\mu_0 = 25$). En $n = 22$ clientes, $\bar{X} = 27.8$, $s = 6.5$. ¿Se superó la meta? ($\alpha = 0.05$, una cola.)

Solución

$H_0 : \mu = 25$ vs. $H_1 : \mu > 25$, gl= 21. $t = \frac{27.8 - 25}{6.5/\sqrt{22}} \approx 2.02$. Crítico $t_{21,0.95} \approx 1.721$. Como $2.02 > 1.721$, **rechazamos** H_0 ; p-valor ≈ 0.028 : sí se superó la meta.

Ejercicio 1.10 (Test de proporción). Históricamente, $p_0 = 40\%$ de los postulantes aprueba un examen de admisión. Con un nuevo curso preparatorio, de $n = 180$ postulantes, 84 aprueban. ¿Cambió la tasa de aprobación? ($\alpha = 0.05$.)

Solución

$\hat{p} = 84/180 \approx 0.467$. $SE = \sqrt{0.40 \cdot 0.60/180} \approx 0.0365$; $z = \frac{0.467 - 0.40}{0.0365} \approx 1.83$. Como $1.83 < 1.96$, **no rechazamos** (p-valor ≈ 0.068): no hay evidencia suficiente de cambio.

Ejercicio 1.11 (Dos muestras independientes, t pooled). Se compara el puntaje de dos secciones de un curso (grupos distintos de estudiantes): Sección A ($n_1 = 12$): $\bar{X}_1 = 68.4$, $s_1 = 7.2$. Sección B ($n_2 = 14$): $\bar{X}_2 = 64.1$, $s_2 = 6.8$ (varianzas similares, usar pooled). ¿Difieren los puntajes medios? ($\alpha = 0.05$.)

Solución

$s_p^2 = \frac{11(7.2)^2 + 13(6.8)^2}{24} \approx 48.81$, $s_p \approx 6.99$; $SE = 6.99\sqrt{1/12 + 1/14} \approx 2.75$; $t = \frac{68.4 - 64.1}{2.75} \approx 1.56$, gl= 24, crítico $t_{24,0.975} \approx 2.064$. Como $1.56 < 2.064$, **no rechazamos** (p-valor ≈ 0.131).

Ejercicio 1.12 (Dos muestras pareadas). Se mide la ansiedad (escala 0–20) de 12 pacientes antes y después de una terapia (mismos pacientes). Las diferencias (después–antes) son: 3, 5, –1, 4, 2, 6, 0, 3, 5, 4, 2, 3. ¿Bajó la ansiedad? ($\alpha = 0.05$, una cola.)

Solución

$\bar{d} = 3.0$, $s_d \approx 2.04$; $t = \frac{3.0}{2.04/\sqrt{12}} \approx 5.08$, gl= 11, crítico $t_{11,0.95} \approx 1.796$. Como $5.08 \gg 1.796$, **rechazamos** H_0 (p-valor ≈ 0.0004): la ansiedad sí bajó.

Ejercicio 1.13 (Dos proporciones independientes). En una encuesta, 64 de 320 jóvenes ($\hat{p}_1 = 0.20$) y 88 de 300 adultos ($\hat{p}_2 = 0.293$) declaran usar cierta red social a diario. ¿Difiere el uso entre grupos? ($\alpha = 0.05$.)

Solución

$\hat{p}_{\text{pool}} = \frac{64 + 88}{320 + 300} \approx 0.245$; $SE \approx 0.0346$; $z = \frac{0.20 - 0.293}{0.0346} \approx -2.70$. Como $|-2.70| = 2.70 > 1.96$, **rechazamos** H_0 (p-valor ≈ 0.007): el uso sí difiere entre grupos.

Ejercicio 1.14 (Chi-cuadrado, bondad de ajuste). Un fabricante de dulces afirma que sus 4 sabores se producen en proporciones iguales. En una muestra de 180 dulces: 42, 58, 35, 45. ¿Coincide con lo afirmado? ($\alpha = 0.05$.)

Solución

Esperado bajo H_0 : 45 cada sabor. $\chi^2 = \sum (O-45)^2/45 \approx 6.18$, gl= 3, crítico $\chi^2_{3,0.95} \approx 7.81$. Como $6.18 < 7.81$, **no rechazamos** (p-valor ≈ 0.103): no hay evidencia de que las proporciones difieran.

Ejercicio 1.15 (Chi-cuadrado, independencia). Se cruza género (fila) con preferencia de un producto nuevo vs. el de siempre (columna): mujeres 54 prefieren el nuevo y 26 el de siempre; hombres 38 prefieren el nuevo y 42 el de siempre. ¿Está asociada la preferencia con el género? ($\alpha = 0.05$.)

Solución

Tabla 2×2 con totales 80 (mujeres), 80 (hombres), 92 (nuevo), 68 (de siempre); esperados = 46 o 34 en cada celda. $\chi^2 \approx 6.55$, gl= 1, crítico $\chi^2_{1,0.95} \approx 3.84$. Como $6.55 > 3.84$, **rechazamos** H_0 (p-valor ≈ 0.010): la preferencia sí está asociada al género.

Ejercicio 1.16 (Reconciliar valor crítico y p-valor). Con $\mu_0 = 100$, $\sigma = 20$ (conocida), $n = 49$, $\bar{X} = 105$: (a) calcula z_{obs} ; (b) compara contra el crítico de dos colas y decide; (c) calcula el p-valor de dos colas y decide; (d) verifica que ambos caminos dan la misma conclusión.

Solución

(a) $z = \frac{105 - 100}{20/\sqrt{49}} = \frac{5}{2.857} \approx 1.75$. (b) $|1.75| < 1.96 \Rightarrow$ no rechazamos (Camino A). (c) p-valor = $2(1 - \Phi(1.75)) \approx 0.080 > 0.05 \Rightarrow$ no rechazamos (Camino B). (d) Misma conclusión por ambos caminos, como debe ser: $|z_{\text{obs}}| < z_{\text{crít}} \iff \text{p-valor} > \alpha$.

Nivel 3 – Aplicación

Combina dos o más ideas en contexto realista; consolida lo aprendido.

Ejercicio 1.17 (RCT: elegir entre pooled y Welch). En un experimento aleatorizado, el grupo tratamiento ($n_1 = 20$) tiene $\bar{X}_1 = 72.5$, $s_1 = 9.0$; el grupo control ($n_2 = 22$) tiene $\bar{X}_2 = 68.0$, $s_2 = 9.5$. ¿El tratamiento *augmentó* el puntaje? ($\alpha = 0.05$, una cola.)

- (a) Justifica por qué aquí es razonable usar t **pooled** en vez de Welch.
- (b) Realiza la prueba completa.

Solución

(a) Las varianzas muestrales (9.0 y 9.5) son parecidas, y el diseño es independiente (sujetos distintos, asignados al azar) $\Rightarrow$ pooled es razonable. (b) $s_p^2 \approx 85.86$, $SE \approx 2.86$; $t = \frac{72.5 - 68.0}{2.86} \approx 1.57$, $gl = 40$, crítico $t_{40,0.95} \approx 1.684$. Como $1.57 < 1.684$, **no rechazamos** (p-valor ≈ 0.062): el efecto no alcanza significancia al 5% (aunque está cerca — posible tema de potencia).

Ejercicio 1.18 (Los mismos datos, dos diseños: antes/después vs. independiente). Se mide el desempeño (escala 0–100) de 12 trabajadores **antes y después** de una capacitación (mismos trabajadores):

Antes	30	55	42	68	50	60	35	65	40	58	45	62
Después	34	59	45	73	54	63	40	69	44	61	50	65

- Analiza correctamente (pareado, pues son los mismos trabajadores). ¿Mejóro el desempeño? ($\alpha = 0.05$, una cola.)
- Ahora trata los 24 números como si fueran dos grupos independientes de 12. ¿Qué concluyes?
- ¿Por qué cambia tanto la conclusión?

Solución

- (a) $\bar{d} = 3.917$, $s_d \approx 0.793$; $t = \frac{3.917}{0.793/\sqrt{12}} \approx 17.11$, $gl = 11$, crítico ≈ 1.796 . $17.11 \gg 1.796 \Rightarrow$ **rechazamos con contundencia** (p-valor $\approx 1.4 \times 10^{-9}$): la capacitación sí mejoró el desempeño.
- (b) Como independientes: $\bar{X}_1 = 50.83$, $\bar{X}_2 = 54.75$, $s_1 \approx 12.39$, $s_2 \approx 12.24$; $t \approx 0.78$, $gl = 22$, crítico ≈ 1.717 . Como $0.78 < 1.717$, **no rechazaríamos** (p-valor ≈ 0.222).
- (c) Al ignorar el pareo, la enorme variabilidad *entre* trabajadores (unos parten en 30, otros en 68) domina y esconde el efecto real, chico pero consistente, de la capacitación. El diseño correcto (pareado) elimina esa variabilidad y revela el efecto con claridad.

Ejercicio 1.19 (Encuesta política: bondad de ajuste). Históricamente, 4 partidos se repartían 30%, 25%, 20% y 25% de la preferencia. En una encuesta reciente ($n = 350$): 88, 102, 75, 85. ¿Cambió la distribución de preferencias? ($\alpha = 0.05$.)

Solución

Esperados: 105, 87.5, 70, 87.5. $\chi^2 \approx 5.58$, $gl = 3$, crítico ≈ 7.81 . Como $5.58 < 7.81$, **no rechazamos** (p-valor ≈ 0.134): no hay evidencia suficiente de cambio.

Ejercicio 1.20 (Tabla 3×2 : nivel educativo y empleo formal).

	Empleo formal	Empleo informal
Primaria	25	75
Secundaria	55	45
Superior	70	30

¿Está asociado el nivel educativo con el tipo de empleo? ($\alpha = 0.05$.)

Solución

Todos los totales de fila suman 100; totales de columna 150 y 150; esperados = 50 en cada celda. $\chi^2 = 42.0$, $gl = (3 - 1)(2 - 1) = 2$, crítico $\chi^2_{2,0.95} \approx 5.99$. Como $42.0 \gg 5.99$, **rechazamos** H_0 (p-valor $\approx 7.6 \times 10^{-10}$): sí hay asociación.

Ejercicio 1.21 (F -test: variabilidad de dos líneas de producción). Línea 1 ($n_1 = 8$): $s_1 = 3.2$. Línea 2 ($n_2 = 13$): $s_2 = 1.8$. ¿Son igual de variables? ($\alpha = 0.05$.)

Solución

$F = \frac{3.2^2}{1.8^2} \approx 3.16$, $gl = (7, 12)$, crítico $F_{7,12,0.95} \approx 2.91$. Como $3.16 > 2.91$, **rechazamos** H_0 (p-valor ≈ 0.039): la línea 1 es significativamente más variable.

Ejercicio 1.22 (Tipo I/II en tamizaje médico, versión numérica). Un test de tamizaje declara “positivo” si un biomarcador supera un umbral k . Bajo ausencia de enfermedad, el biomarcador es $\mathcal{N}(100, 20^2/25)$ (es decir, $\mu_0 = 100$, $\sigma = 20$, $n = 25$); bajo enfermedad, $\mathcal{N}(115, 20^2/25)$. Con $\alpha = 0.05$ (una cola derecha):

- Calcula el umbral k .
- Calcula β y la potencia (probabilidad de detectar la enfermedad si está presente).

Solución

(a) $k = 100 + 1.645 \cdot \frac{20}{\sqrt{25}} = 100 + 6.58 = 106.58$.

(b) $\beta = \Phi\left(\frac{106.58 - 115}{20/\sqrt{25}}\right) = \Phi(-2.105) \approx 0.018$; potencia = $1 - \beta \approx 0.982$. Con este diseño, el test detecta correctamente $\approx 98\%$ de los casos enfermos.

Ejercicio 1.23 (Elegir el test correcto: seis escenarios mezclados). Para cada escenario, di qué prueba usarías (test, y si aplica, pareado/independiente):

- Comparar el gasto promedio de dos ciudades (grupos distintos de hogares).
- Ver si el mismo grupo de estudiantes mejora su nota tras un taller.
- Ver si el color favorito está asociado con la región de residencia.

- (d) Probar si la proporción de hogares con internet superó el 60 %.
- (e) Comparar la variabilidad de tiempos de entrega de dos proveedores.
- (f) Probar si el ingreso medio de un distrito *difiere* de S/1500, con $n = 15$ y σ desconocida.

Solución

(a) t de dos muestras independientes (pooled o Welch según varianzas). (b) t pareado. (c) χ^2 de independencia. (d) z de una proporción, una cola. (e) F -test. (f) t de una muestra, dos colas.

Nivel 4 – Reto

Difícil: integra varios temas en un mismo problema.

Ejercicio 1.24 (Derivar el tamaño de muestra para una potencia objetivo). Se quiere detectar un efecto $\Delta = 4$ (sobre una variable con $\sigma = 10$) con potencia objetivo $1 - \beta = 0.90$ y $\alpha = 0.05$ (una cola).

- (a) Escribe la fórmula de n y los valores de $z_{1-\alpha}$ y $z_{1-\beta}$ que necesitas.
- (b) Calcula n (redondeando hacia arriba).

Solución

$$(a) \ n = \frac{\sigma^2(z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2}, \text{ con } z_{1-\alpha} = z_{0.95} = 1.645 \text{ y } z_{1-\beta} = z_{0.90} = 1.282.$$

$$(b) \ n = \frac{10^2(1.645 + 1.282)^2}{4^2} = \frac{100 \times 8.566}{16} \approx 53.5 \Rightarrow n = 54.$$

Ejercicio 1.25 (Welch vs. pooled con varianzas muy distintas). Grupo 1 ($n_1 = 8$): $\bar{X}_1 = 55.0$, $s_1 = 15.0$ (muy disperso). Grupo 2 ($n_2 = 30$): $\bar{X}_2 = 50.5$, $s_2 = 4.0$ (muy estable).

- (a) Calcula el error estándar y el estadístico con Welch.
- (b) Calcula el error estándar y el estadístico con pooled (aunque no corresponde).
- (c) Explica por qué pooled da una impresión engañosa aquí.

Solución

- (a) Welch: $SE = \sqrt{15^2/8 + 4^2/30} \approx 5.35$; $t \approx 0.84$, gl (Welch) ≈ 7.3 .
- (b) Pooled: $s_p^2 \approx 56.64$; $SE \approx 2.99$; $t \approx 1.50$, gl = 36.
- (c) Pooled *promedia* las dos varianzas ponderando por tamaño de muestra, pero el grupo

1 (chico y muy disperso) aporta poca información real; al pooled tratarlo como si su varianza fuera representativa junto con el grupo grande y estable, el SE resultante es artificialmente **más chico** (más “confianza” de la que hay), y el gl usado (36, mucho mayor al gl Welch ≈ 7.3) exagera la precisión. Welch es más honesto aquí.

Ejercicio 1.26 (La trampa de los conteos esperados chicos). En una encuesta sobre un evento raro, se clasifican 50 respuestas en 4 categorías con proporciones históricas 5 %, 10 %, 5 %, 80 %; observado: 2, 3, 1, 44.

- (a) Calcula los conteos esperados de cada categoría.
- (b) Calcula χ^2 y el p-valor.
- (c) ¿Por qué este resultado no es confiable, aunque el cálculo sea correcto?

Solución

- (a) Esperados: 2.5, 5.0, 2.5, 40.0.
- (b) $\chi^2 \approx 2.20$, gl= 3, p-valor ≈ 0.532 .
- (c) Tres de las cuatro celdas esperadas son menores a 5 (regla práctica mínima); la aproximación χ^2 deja de ser confiable con conteos esperados tan chicos. Habría que agrupar categorías (p. ej. fusionar las tres categorías raras en “otra”) o usar el test exacto de Fisher.

Ejercicio 1.27 (F -test como preludio a la regresión). Explica, en tus propias palabras, por qué el estadístico F (razón de dos varianzas) reaparecerá más adelante como la prueba de significancia **global** de un modelo de regresión. (Pista: piensa en “varianza explicada por el modelo” vs. “varianza que sobra sin explicar”).

Solución

En regresión, se compara la varianza que el modelo *explica* (varianza de los valores ajustados) contra la varianza *residual* (lo que el modelo no explica). Esa razón sigue –también– una distribución F bajo la hipótesis nula de que el modelo no aporta nada; es exactamente la misma lógica de “comparar dos varianzas” que vimos aquí, aplicada a un contexto distinto.

Ejercicio 1.28 (Dualidad IC $\leftrightarrow$ prueba, para dos muestras). Para el Ejercicio de dos muestras independientes pooled (Secciones A y B, Núcleo), construye el IC del 95 % para $\mu_1 - \mu_2$ y verifica que la conclusión coincide con la prueba de hipótesis de dos colas.

Solución

IC: $(\bar{X}_1 - \bar{X}_2) \pm t_{24,0.975} \cdot \text{SE} = 4.3 \pm 2.064(2.75) \approx 4.3 \pm 5.68 = [-1.38, 9.98]$. Como el intervalo **incluye** el 0, coincide con no rechazar $H_0 : \mu_1 = \mu_2$ en la prueba de dos colas (que tampoco rechazaba). La dualidad se mantiene: IC contiene 0 $\iff$ no se rechaza

al nivel correspondiente.

Ejercicio 1.29 (Separar tres preguntas mezcladas). Un informe dice: “Queremos saber si el ingreso medio de los beneficiarios *superó* el de los no beneficiarios tras el programa (mismos hogares, medidos antes y después), si la distribución de nivel de pobreza cambió respecto a la histórica, y si la variabilidad del ingreso es distinta entre zona urbana y rural.” Identifica las tres preguntas y la prueba de cada una.

Solución

(1) “Superó... mismos hogares, antes y después” $\rightarrow t$ **pareado**, una cola. (2) “Distribución de nivel de pobreza cambió respecto a la histórica” $\rightarrow \chi^2$ de **bondad de ajuste**. (3) “Variabilidad... distinta entre zona urbana y rural” $\rightarrow F$ -**test**.

Nivel 5 – Reto+

Muy difícil / fuera del scope típico: demostraciones y casos límite.

Ejercicio 1.30 (Derivación: $z^2 \sim \chi_1^2$). Sea $Z \sim \mathcal{N}(0, 1)$. Demuestra que Z^2 sigue una distribución χ^2 con 1 grado de libertad, usando la función generadora de momentos (MGF) o un cambio de variable directo sobre la densidad de Z .

Solución

Vía cambio de variable: sea $W = Z^2$. Para $w > 0$, $F_W(w) = \mathbb{P}(Z^2 \leq w) = \mathbb{P}(-\sqrt{w} \leq Z \leq \sqrt{w}) = 2\Phi(\sqrt{w}) - 1$. Derivando, $f_W(w) = 2\phi(\sqrt{w}) \cdot \frac{1}{2\sqrt{w}} = \frac{1}{\sqrt{2\pi w}} e^{-w/2}$, que es exactamente la densidad χ_1^2 (caso particular de χ_k^2 con $k = 1$: $f(w) = \frac{1}{2^{1/2}\Gamma(1/2)} w^{-1/2} e^{-w/2}$, y $\Gamma(1/2) = \sqrt{\pi}$). Por lo tanto $Z^2 \sim \chi_1^2$. (Esto es un caso particular del teorema de Wilks, ya visto en el ejercicio Reto+ de razón de verosimilitud de la Clase 6.)

Ejercicio 1.31 (El t pareado como caso particular del t de una muestra). Muestra formalmente que el test t pareado es un test t de una muestra aplicado a las diferencias $d_i = X_i - Y_i$.

Solución

Sea $d_i = X_i - Y_i$, $i = 1, \dots, n$, con $\bar{d} = \frac{1}{n} \sum d_i$ y $s_d^2 = \frac{1}{n-1} \sum (d_i - \bar{d})^2$. La hipótesis “ $\mu_X = \mu_Y$ ” es equivalente a “ $\mu_d = \mathbb{E}[d_i] = 0$ ”. El estadístico $t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$ es *exactamente* el estadístico t de una muestra (Sección 5) aplicado a la variable d con $\mu_0 = 0$, y sigue t_{n-1} bajo H_0 por el mismo argumento (razón entre una Normal estándar y la raíz de una χ_{n-1}^2 escalada, independientes). No hay nada nuevo que derivar: es una aplicación directa.

Ejercicio 1.32 (Bootstrap para dos muestras, sin supuesto de normalidad). Describe, conceptualmente, cómo construirías una prueba de diferencia de medianas entre dos grupos independientes usando **bootstrap** (remuestreo), sin asumir normalidad. ¿Qué estadístico remuestrearías y cómo obtendrías un p-valor?

Solución

(1) Calcula la diferencia observada de medianas $\hat{\Delta} = \text{med}(X) - \text{med}(Y)$. (2) Bajo H_0 (misma mediana), combina ambas muestras en un solo conjunto de $n_1 + n_2$ observaciones. (3) Remuestrea (con reemplazo, o por permutación sin reemplazo) dos grupos del mismo tamaño n_1, n_2 desde el conjunto combinado, y calcula la diferencia de medianas remuestreada $\hat{\Delta}^{(b)}$, repitiendo $B = 10\,000$ veces. (4) El p-valor (dos colas) es la proporción de $|\hat{\Delta}^{(b)}| \geq |\hat{\Delta}|$. Esto evita asumir normalidad o una fórmula cerrada para la varianza de la mediana.

Ejercicio 1.33 (Pooled y Welch coinciden cuando $n_1 = n_2$ y $s_1 = s_2$). Muestra algebraicamente que, si $n_1 = n_2 = n$ y $s_1 = s_2 = s$, el estadístico t pooled y el estadístico t de Welch son idénticos (y por lo tanto también coinciden sus grados de libertad).

Solución

Con $n_1 = n_2 = n$, $s_1 = s_2 = s$: $s_p^2 = \frac{(n-1)s^2 + (n-1)s^2}{2n-2} = s^2$, así que $\text{SE}_{\text{pooled}} = s\sqrt{1/n + 1/n} = s\sqrt{2/n}$. Para Welch, $\text{SE}_{\text{Welch}} = \sqrt{s^2/n + s^2/n} = s\sqrt{2/n}$: idéntico. Para los grados de libertad, pooled da $\text{gl} = 2n - 2$; Welch–Satterthwaite da

$$\text{gl} = \frac{(s^2/n + s^2/n)^2}{(s^2/n)^2/(n-1) + (s^2/n)^2/(n-1)} = \frac{(2s^2/n)^2}{2(s^2/n)^2/(n-1)} = (n-1) \cdot 2 = 2n - 2,$$

también idéntico. Con grupos balanceados y de igual varianza, la distinción pooled/Welch desaparece por completo.

Fuentes y atribución

Material de repaso **basado en MIT 14.310x – Data Analysis for Social Scientists** (Esther Duflo & Sara Fisher Ellison), MIT OpenCourseWare, licencia **CC BY-NC-SA 4.0** (uso no comercial, con atribución, compartir bajo la misma licencia), y en las fuentes de la bibliografía. Los enunciados se basan en el material del curso y en diversas fuentes abiertas; cuando un problema se adapta de una fuente específica, se cita en el enunciado.

Bibliografía.

- Duflo, E. & Ellison, S. F. *14.310x: Data Analysis for Social Scientists*. MIT OpenCourseWare (slides, lecture notes y problem sets). ocw.mit.edu.
- Larsen, R. J. & Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 6.^a ed., Pearson, 2018.
- DeGroot, M. H. & Schervish, M. J. *Probability and Statistics*, 4.^a ed., Pearson, 2012.
- Blitzstein, J. & Hwang, J. *Introduction to Probability* (Harvard Stat 110); W. Chen, *Probability Cheatsheet*.
- Diez, D., Çetinkaya-Rundel, M. & Barr, C. *OpenIntro Statistics*. openintro.org.
- Materiales abiertos de la comunidad del curso en GitHub: [gevorg-hunanyan/probability](https://github.com/gevorg-hunanyan/probability); [hanmochen/Probability-Theory-2020](https://github.com/hanmochen/Probability-Theory-2020); [mynameisjanus/14310xDataAnalysis](https://github.com/mynameisjanus/14310xDataAnalysis).

creativecommons.org/licenses/by-nc-sa/4.0