

14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

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

Clase 6

Estimadores, Intervalos de Confianza y Pruebas de Hipótesis

Ejercicios (5 niveles)

Módulos oficiales del MicroMaster:

M6 – Assessing and Deriving Estimators, Confidence Intervals and Hypothesis Testing

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.

Indicaciones

Resuelve los ejercicios de tu nivel; si terminas, sube al siguiente (ver la leyenda de arriba). Recuerda: el *sesgo* es $\text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$ y $\text{MSE}(\hat{\theta}) = \text{bias}^2 + \text{Var}(\hat{\theta})$; el IC del 95 % para la media con σ conocida es $\bar{x} \pm z_{0.975} \sigma / \sqrt{n}$ (con $z_{0.975} = 1.96$, $z_{0.95} = 1.645$, $z_{0.995} = 2.576$); el estadístico de prueba para una media es $z = (\bar{x} - \mu_0) / (\sigma / \sqrt{n})$. Verifica siempre unidades, parámetros y el sentido de H_0 frente a H_1 (una o dos colas) antes de decidir.

1. Problemas

Nivel 1 – Fundamentos (warm-up)

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

Ejercicio 1.1 (¿El estimador es insesgado?). Sea $X_1, \dots, X_n$ una muestra aleatoria i.i.d. con media poblacional μ y varianza σ^2 . Se proponen dos estimadores de μ : $\hat{\mu}_1 = \bar{X}$ (la media muestral) y $\hat{\mu}_2 = X_1$ (la primera observación).

- (a) Calcula $\mathbb{E}[\hat{\mu}_1]$ y $\mathbb{E}[\hat{\mu}_2]$. ¿Son insesgados para μ ?
- (b) El *sesgo* (bias) de un estimador $\hat{\theta}$ es $\text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$. Da el sesgo de cada uno.

Ejercicio 1.2 (Calcular el sesgo de un estimador propuesto). Sea $X_1, \dots, X_n$ i.i.d. con media μ . Un analista propone estimar μ con $\hat{\mu}_3 = \bar{X} + 2$ (suma 2 “por si acaso”).

- (a) Calcula $\mathbb{E}[\hat{\mu}_3]$ usando que $\mathbb{E}[\bar{X}] = \mu$.
- (b) Da el sesgo $\text{bias}(\hat{\mu}_3) = \mathbb{E}[\hat{\mu}_3] - \mu$. ¿Es insesgado?

Ejercicio 1.3 (Error estándar de la media muestral). El gasto mensual en alimentos en una población tiene desviación estándar $\sigma = 240$ soles (conocida). Se toma una muestra aleatoria de $n = 64$ hogares y se calcula $\bar{X}$.

- (a) Calcula el error estándar $SD(\bar{X}) = \sigma/\sqrt{n}$.
- (b) Si n sube a 144, ¿cuánto vale el error estándar?

Ejercicio 1.4 (Construir un IC del 95 % paso a paso (σ conocida)). El peso al nacer en una maternidad tiene desviación estándar poblacional conocida $\sigma = 500$ g. En una muestra de $n = 25$ recién nacidos la media es $\bar{x} = 3\,200$ g. El IC del 95 % para μ es $\bar{x} \pm z_{0.975} \frac{\sigma}{\sqrt{n}}$, con $z_{0.975} = 1.96$.

- (a) Calcula el error estándar $\sigma/\sqrt{n}$.
- (b) Calcula el margen de error $z_{0.975} \sigma/\sqrt{n}$.
- (c) Escribe el intervalo de confianza del 95 % para μ .

Ejercicio 1.5 (IC del 95 % para la media con σ conocida). Una prueba estandarizada tiene desviación estándar poblacional conocida $\sigma = 15$ puntos. En una muestra de $n = 100$ estudiantes la media es $\bar{x} = 72$. El IC del 95 % para μ es $\bar{x} \pm z_{0.975} \frac{\sigma}{\sqrt{n}}$, con $z_{0.975} = 1.96$.

- (a) Calcula el error estándar $\sigma/\sqrt{n}$ y el margen de error $z_{0.975} \sigma/\sqrt{n}$.
- (b) Escribe el intervalo de confianza del 95 %.

Ejercicio 1.6 (Identificar H_0 y H_1 de una frase). Para cada situación plantea la hipótesis nula H_0 y la alternativa H_1 en términos del parámetro indicado.

- (a) Se afirma que el tiempo medio de trámite es $\mu_0 = 30$ min; quieres ver si *difiere* de 30.
- (b) Una ONG sostiene que su programa eleva el ingreso medio por encima de $\mu_0 = 1\,500$ soles.
- (c) Un funcionario afirma que la proporción de aprobación es $p_0 = 0.5$; sospechas que es *menor*.

Ejercicio 1.7 (Estimación puntual de μ y de p). En una encuesta a $n = 200$ personas, 130 aprueban una política y la edad media observada es $\bar{x} = 41$ años.

- (a) Da la estimación puntual de la proporción poblacional p que aprueba.
- (b) Da la estimación puntual de la media poblacional de edad μ .

Ejercicio 1.8 (Leer un valor crítico). Trabajas con el estadístico Z bajo H_0 , que es $\mathcal{N}(0, 1)$.

- (a) Para una prueba de *dos colas* con $\alpha = 0.05$, ¿cuál es el valor crítico $z_{1-\alpha/2}$ y la regla de rechazo?
- (b) Para una prueba de *una cola* (derecha) con $\alpha = 0.05$, ¿cuál es el valor crítico $z_{1-\alpha}$ y la regla de rechazo?

Nivel 2 – Núcleo (core)

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

Ejercicio 1.9 (MSE = sesgo² + varianza). El error cuadrático medio de un estimador es $\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \text{bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$. Un estimador del ingreso medio tiene sesgo $\text{bias}(\hat{\theta}) = 4$ soles y varianza $\text{Var}(\hat{\theta}) = 9$ soles².

- (a) Calcula su MSE.
- (b) Otro estimador es insesgado pero tiene varianza 20. ¿Cuál tiene menor MSE?

Ejercicio 1.10 (Interpretar correctamente un IC). Un equipo reporta un IC del 95 % para la proporción de hogares con internet: $[0.58, 0.66]$. Para cada afirmación indica si es **correcta** o **incorrecta**.

- (a) “Hay 95 % de probabilidad de que el verdadero p esté entre 0.58 y 0.66.”
- (b) “Si repitiéramos el muestreo muchas veces, cerca del 95 % de los intervalos así contruidos contendrían al verdadero p .”
- (c) “El 95 % de los hogares de la muestra tiene internet.”

Ejercicio 1.11 (Nivel de confianza y ancho del IC). Con $\bar{x} = 50$, $\sigma = 10$ y $n = 100$ (así $\sigma/\sqrt{n} = 1$), se calculan dos intervalos para μ : uno al 90 % ($z = 1.645$) y otro al 99 % ($z = 2.576$).

- (a) Escribe ambos intervalos.
- (b) ¿Cuál es más ancho y por qué?

Ejercicio 1.12 (Decisión con un valor crítico). Para probar $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$ se calcula el estadístico $z = 2.3$. Usas $\alpha = 0.05$ (dos colas), con valor crítico $z_{0.975} = 1.96$.

- (a) ¿Se rechaza H_0 ? Justifica.
- (b) Si el estadístico hubiera sido $z = 1.2$, ¿cambiaría la decisión?

Ejercicio 1.13 (IC para la media con σ desconocida (t de Student)). En $n = 16$ municipios se mide el gasto per cápita en salud: $\bar{x} = 320$ soles y desviación estándar *muestral* $s = 40$ soles. Como σ es desconocida, se usa $\bar{x} \pm t_{n-1, 0.975} \frac{s}{\sqrt{n}}$. Con 15 grados de libertad, $t_{15, 0.975} = 2.131$.

- (a) Calcula el error estándar estimado $s/\sqrt{n}$ y el margen de error.
- (b) Escribe el IC del 95 % y compáralo con el que daría la normal ($z = 1.96$).

Ejercicio 1.14 (IC para una proporción). En una encuesta a $n = 400$ votantes, $\hat{p} = 0.45$ apoya a una candidata. El IC del 95 % para p usa $\hat{p} \pm z_{0.975} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ con $z_{0.975} = 1.96$.

(a) Calcula el error estándar $\sqrt{\hat{p}(1-\hat{p})/n}$ y el margen de error.

(b) Escribe el intervalo. ¿Es plausible que el verdadero p sea 0.5?

Ejercicio 1.15 (Prueba de hipótesis sobre una media (estadístico z)). Se afirma que el tiempo medio de espera en una posta es $\mu_0 = 20$ min. Con $\sigma = 6$ min conocida, una muestra de $n = 36$ da $\bar{x} = 22$ min. Prueba $H_0 : \mu = 20$ vs. $H_1 : \mu \neq 20$ con $\alpha = 0.05$.

(a) Calcula el estadístico $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.

(b) Compara con $z_{0.975} = 1.96$ y concluye.

Ejercicio 1.16 (p -valor y su interpretación). En la prueba anterior ($H_1 : \mu \neq 20$) se obtuvo $z = 2$.

(a) Calcula el p -valor de dos colas, $p = 2\mathbb{P}(Z > 2)$. (Usa $\mathbb{P}(Z > 2) \approx 0.0228$.)

(b) Interpreta el p -valor y compáralo con $\alpha = 0.05$.

Ejercicio 1.17 (Error tipo I vs. error tipo II). Se prueba si un programa social funciona (H_0 : “no tiene efecto”; H_1 : “sí tiene efecto”).

(a) Describe en este contexto el error tipo I y el error tipo II.

(b) Nombra α , β y la potencia, y di qué mide cada uno.

Ejercicio 1.18 (Prueba de una cola sobre la media). Una ONG afirma que su programa eleva el ingreso medio por encima de $\mu_0 = 1\,200$ soles. Con $\sigma = 300$ conocida, una muestra de $n = 100$ beneficiarios da $\bar{x} = 1\,260$. Prueba $H_0 : \mu = 1\,200$ vs. $H_1 : \mu > 1\,200$ con $\alpha = 0.05$.

(a) Calcula el estadístico z .

(b) Compáralo con $z_{0.95} = 1.645$ y concluye.

Nivel 3 – Aplicación

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

Ejercicio 1.19 (Efecto del tamaño de muestra en el ancho del IC). Con $\sigma = 20$ conocida y $\bar{x}$ fijo, se construye un IC del 95 % para μ ($z = 1.96$).

(1) Da el margen de error $m = z\sigma/\sqrt{n}$ para $n = 25$ y para $n = 100$.

(2) ¿En qué factor cambia el ancho al pasar de $n = 25$ a $n = 100$?

Ejercicio 1.20 (Tamaño de muestra para un margen de error dado). Quieres estimar la proporción p de hogares con acceso a agua potable con un margen de error de a lo sumo 0.03 al 95 % ($z = 1.96$). Sin información previa usas el caso más conservador $p = 0.5$ (que maximiza $p(1-p) = 0.25$).

(1) Plantea $z\sqrt{p(1-p)/n} \leq 0.03$ y despeja n .

(2) ¿Cuántos hogares se necesitan?

Ejercicio 1.21 (IC del 95 % y dualidad con una prueba de dos colas). Con $\sigma = 12$ conocida, una muestra de $n = 144$ estudiantes da $\bar{x} = 64$. Además se quiere contrastar si μ podría ser 66.

(1) Construye el IC del 95 % para μ ($z = 1.96$).

(2) Usando el IC, ¿se rechazaría $H_0 : \mu = 66$ vs. $H_1 : \mu \neq 66$ al 5 %?

Ejercicio 1.22 (Decisión por p -valor en una cola). Se prueba $H_0 : p = 0.30$ vs. $H_1 : p > 0.30$ (proporción de jóvenes que estudian y trabajan). El estadístico calculado es $z = 1.75$.

(1) Calcula el p -valor de una cola, $p = \mathbb{P}(Z > 1.75)$. (Usa $\mathbb{P}(Z > 1.75) \approx 0.0401$.)

(2) Decide con $\alpha = 0.05$ y con $\alpha = 0.01$.

Ejercicio 1.23 (Estimador por momentos y por máxima verosimilitud: Bernoulli). Sea $X_1, \dots, X_n$ i.i.d. Bernoulli(p) (por ejemplo, $X_i = 1$ si el hogar i tiene internet). Recuerda $\mathbb{E}[X_i] = p$ y $f(x | p) = p^x(1-p)^{1-x}$.

(1) **Momentos:** iguala el primer momento poblacional al muestral y despeja $\hat{p}_{\text{MM}}$.

(2) **Máxima verosimilitud:** escribe la log-verosimilitud $\ell(p) = \sum_i [x_i \ln p + (1-x_i) \ln(1-p)]$, deriva, iguala a cero y despeja $\hat{p}_{\text{MLE}}$.

Ejercicio 1.24 (MLE de λ : Poisson y Exponencial). (1) **Poisson.** $X_1, \dots, X_n$ i.i.d. $\text{Pois}(\lambda)$, con $f(x | \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$. Halla $\hat{\lambda}_{\text{MLE}}$.

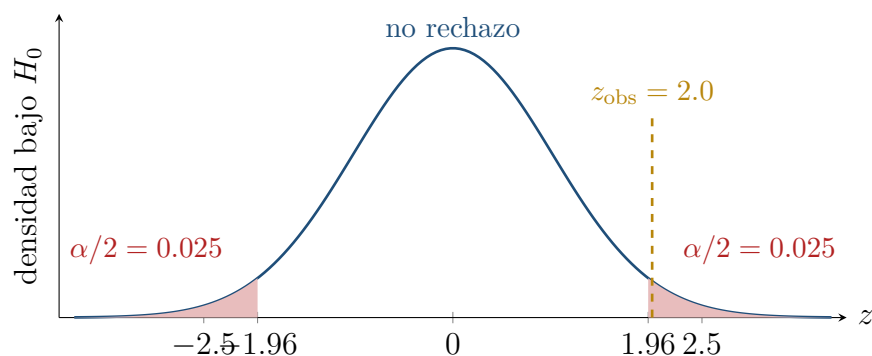
(2) **Exponencial.** $T_1, \dots, T_n$ i.i.d. exponenciales con tasa λ , $f(t | \lambda) = \lambda e^{-\lambda t}$ ($t \geq 0$). Halla $\hat{\lambda}_{\text{MLE}}$.

Ejercicio 1.25 (Consistencia de la media muestral). Sea $\bar{X}$ la media de una muestra i.i.d. con media μ y varianza $\sigma^2 < \infty$. Un estimador $\hat{\theta}_n$ es *consistente* para θ si $\hat{\theta}_n \xrightarrow{p} \theta$ cuando $n \rightarrow \infty$.

(1) Calcula $\text{MSE}(\bar{X})$ y muestra que $\rightarrow 0$ cuando $n \rightarrow \infty$.

(2) Argumenta (vía la desigualdad de Chebyshev / convergencia en media cuadrática) que $\bar{X}$ es consistente para μ .

Ejercicio 1.26 (Región de rechazo y p -valor bajo la Normal). La figura muestra la densidad del estadístico Z bajo H_0 ($Z \sim \mathcal{N}(0, 1)$) en una prueba de *dos colas* con $\alpha = 0.05$. Las colas rojas son la región de rechazo ($|z| > 1.96$); la línea punteada marca un estadístico observado $z_{\text{obs}} = 2.0$.



- (1) ¿Se rechaza H_0 con $z_{\text{obs}} = 2.0$? Justifica con la figura.
- (2) Calcula el p -valor de dos colas asociado a $z_{\text{obs}} = 2.0$. (Usa $\mathbb{P}(Z > 2) \approx 0.0228$.)

Nivel 4 – Reto

Difícil: integra varios temas en un mismo problema.

Ejercicio 1.27 (Comparar dos estimadores por MSE). Para estimar μ con muestra i.i.d. ($\text{Var}(X_i) = \sigma^2$) se comparan $\hat{\mu}_1 = \bar{X}$ (insesgado, $\text{Var} = \sigma^2/n$) y el estimador encogido $\hat{\mu}_2 = \frac{1}{2}\bar{X}$ (sesgado hacia 0).

- (1) Calcula $\text{bias}(\hat{\mu}_2)$ y $\text{Var}(\hat{\mu}_2)$ en términos de μ, σ^2, n .
- (2) Escribe $\text{MSE}(\hat{\mu}_1)$ y $\text{MSE}(\hat{\mu}_2)$. Para $\mu = 4$, $\sigma^2 = 4$, $n = 4$, ¿cuál es menor?

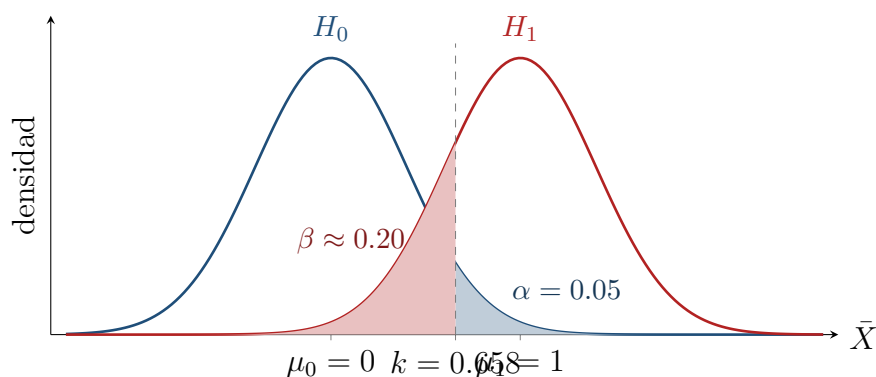
Ejercicio 1.28 (Cálculo de potencia $(1 - \beta)$). Se prueba $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$ con $\sigma = 2$ conocida y $n = 25$, así $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ con $\text{SD}(\bar{X}) = 2/5 = 0.4$. Se usa $\alpha = 0.05$, de modo que la regla rechaza si $\bar{X} > k$ con $k = 0 + 1.645(0.4) = 0.658$.

- (1) Verifica el valor crítico k a partir de $\mathbb{P}(\bar{X} > k \mid \mu = 0) = 0.05$.
- (2) Calcula la potencia si el verdadero valor es $\mu = 1$, es decir $1 - \beta = \mathbb{P}(\bar{X} > k \mid \mu = 1)$. (Usa $\mathbb{P}(Z > -0.855) = \mathbb{P}(Z < 0.855) \approx 0.80$.)

Ejercicio 1.29 (Prueba de dos colas vs. una cola). Se contrasta una media con σ conocida y se obtiene $z = 1.8$.

- (1) Con $H_1 : \mu > \mu_0$ (una cola, $\alpha = 0.05$, $z_{0.95} = 1.645$), ¿se rechaza H_0 ? Da también el p -valor de una cola. (Usa $\mathbb{P}(Z > 1.8) \approx 0.0359$.)
- (2) Con $H_1 : \mu \neq \mu_0$ (dos colas, $\alpha = 0.05$, $z_{0.975} = 1.96$), ¿se rechaza? Da el p -valor de dos colas.

Ejercicio 1.30 (Las dos densidades H_0 y $H_1: \alpha$ y β). La figura muestra la distribución de $\bar{X}$ bajo $H_0 : \mu = 0$ (azul) y bajo $H_1 : \mu = 1$ (rojo), ambas con $\text{SD}(\bar{X}) = 0.4$ (caso $\sigma = 2$, $n = 25$). El umbral de rechazo de una cola es $k = 0.658$. El área azul a la derecha de k es $\alpha = 0.05$; el área roja a la izquierda de k es $\beta \approx 0.20$.

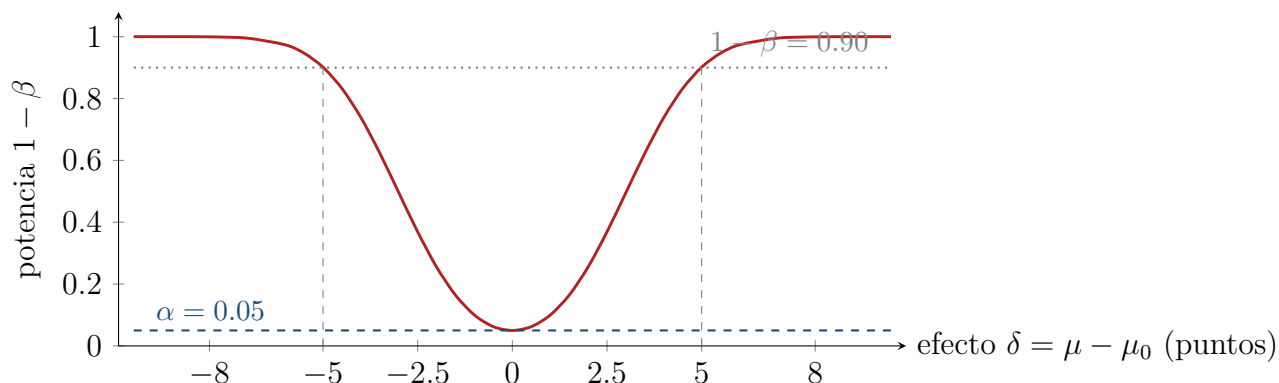


- (1) Identifica en la figura α , β y la potencia $1 - \beta$.
- (2) Si n aumenta, ¿qué les pasa al ancho de las campanas y a β (con α y k ajustados)?

Ejercicio 1.31 (Tamaño de muestra para una potencia objetivo y curva de potencia). Un ministerio quiere evaluar si una intervención cambia el puntaje medio μ de una prueba estandarizada cuya desviación estándar poblacional es $\sigma = 15$ puntos (conocida). Se contrasta $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ con $\alpha = 0.05$ (dos colas, $z_{0.975} = 1.96$). Se desea *detectar* un efecto de tamaño $\delta = 5$ puntos con potencia $1 - \beta = 0.90$. La fórmula aproximada de tamaño de muestra para una prueba de dos colas es

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\delta^2}, \quad z_{0.975} = 1.96, \quad z_{0.90} = 1.28.$$

- (1) Calcula el n necesario y redondéalo hacia arriba.
- (2) Con ese n , el error estándar es $SD(\bar{X}) = \sigma/\sqrt{n}$. Explica por qué la potencia *crece* con $|\delta|$ y se aproxima a α cuando $\delta \rightarrow 0$. Apoya tu argumento en la curva de potencia de la figura.
- (3) Si solo importara detectar un efecto la mitad de grande ($\delta = 2.5$), ¿en qué factor cambiaría el n requerido?



Nivel 5 – Reto+*Muy difícil / fuera del scope típico: demostraciones y casos límite.*

Ejercicio 1.32 (MLE de μ y σ^2 de una Normal: el sesgo del MLE de σ^2). Sea $X_1, \dots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$ con *ambos* parámetros desconocidos. La densidad es $f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$.

- (1) Escribe la log-verosimilitud $\ell(\mu, \sigma^2)$ y deriva los MLE $\hat{\mu}$ y $\hat{\sigma}^2$ resolviendo $\partial\ell/\partial\mu = 0$ y $\partial\ell/\partial\sigma^2 = 0$.
- (2) Demuestra que $\hat{\sigma}^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2$ es *sesgado*: calcula $\mathbb{E}[\hat{\sigma}^2]$ usando $\mathbb{E}\left[\sum_i (X_i - \bar{X})^2\right] = (n-1)\sigma^2$. ¿Cuál es el sesgo y hacia dónde apunta?
- (3) ¿Qué corrección simple lo vuelve insesgado? Comentá por qué el MLE puede ser sesgado pese a ser “óptimo” en el sentido de la verosimilitud.

Ejercicio 1.33 (Comparar dos estimadores de σ^2 por MSE: insesgado vs. encogido). Se quiere estimar la varianza σ^2 de una población $\mathcal{N}(\mu, \sigma^2)$ con $n = 10$ observaciones. Se comparan dos estimadores de la forma $c \sum_i (X_i - \bar{X})^2$:

$$T_1 = S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 \quad (\text{insesgado}), \quad T_2 = \hat{\sigma}_{\text{MLE}}^2 = \frac{1}{n} \sum_i (X_i - \bar{X})^2 \quad (\text{sesgado}).$$

Para la Normal se sabe que $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$.

- (1) Calcula $\text{MSE}(T_1)$ (recuerda que es insesgado).
- (2) Escribe $T_2 = \frac{n-1}{n} S^2$. Halla su sesgo y su varianza en términos de σ^2 , y luego $\text{MSE}(T_2)$. Con $n = 10$ y tomando $\sigma^2 = 1$, ¿cuál estimador tiene menor MSE?
- (3) ¿Qué te dice este resultado sobre preferir siempre el estimador insesgado?

Ejercicio 1.34 (*p*-hacking: pruebas múltiples inflan los falsos positivos). Un equipo de investigación recolecta datos de una encuesta social y, sin una hipótesis previa, prueba m correlaciones distintas, cada una con $\alpha = 0.05$. Supongamos (escenario nulo) que *ninguna* relación es real, y que las m pruebas son independientes.

- (1) Para una sola prueba, ¿cuál es la probabilidad de *no* cometer un falso positivo? ¿Y para las m pruebas (la probabilidad de no tener *ningún* falso positivo)?
- (2) Calcula $\mathbb{P}(\text{al menos un falso positivo}) = 1 - (1 - \alpha)^m$ para $m = 10$ y $m = 20$. Comenta el resultado.

- (3) **Corrección de Bonferroni:** para mantener la probabilidad familiar de error (FWER) en ≤ 0.05 , se usa $\alpha^* = \alpha/m$ en cada prueba. Calcula α^* para $m = 20$ y verifica que $1 - (1 - \alpha^*)^{20} \approx 0.05$. ¿Qué se sacrifica al aplicar la corrección?

Ejercicio 1.35 (Prueba de la razón de verosimilitud (introducción)). Sea $X_1, \dots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$ con σ^2 conocida. Se contrasta el nulo simple $H_0 : \mu = \mu_0$ contra el alternativo compuesto $H_1 : \mu \neq \mu_0$. El estadístico de la razón de verosimilitud (generalizada) es

$$\Lambda = \frac{\sup_{\mu} L(\mu)}{L(\mu_0)}, \quad L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

- (1) El supremo de $L(\mu)$ se alcanza en el MLE $\hat{\mu} = \bar{x}$. Demuestra que

$$2 \ln \Lambda = \frac{1}{\sigma^2} \left[\sum_i (x_i - \mu_0)^2 - \sum_i (x_i - \bar{x})^2 \right] = \frac{n(\bar{x} - \mu_0)^2}{\sigma^2}.$$

(Usa la identidad $\sum_i (x_i - \mu_0)^2 = \sum_i (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2$.)

- (2) Concluye que $2 \ln \Lambda = z^2$, donde $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ es el estadístico z usual. Así, “rechazar cuando Λ es grande” equivale a “rechazar cuando $|z|$ es grande”. ¿Qué distribución tiene $2 \ln \Lambda$ bajo H_0 ?
- (3) Si $\sigma = 4$, $n = 16$, $\mu_0 = 10$ y $\bar{x} = 12$, calcula $2 \ln \Lambda$ y decide al 5 % usando que el valor crítico de una χ_1^2 es 3.84 ($= 1.96^2$).

Ejercicio 1.36 (Bootstrap conceptual: IC sin fórmula cerrada). Un economista tiene una muestra de $n = 200$ ingresos y quiere un IC del 95 % para la *mediana* poblacional, para la cual no hay una fórmula simple del error estándar (la mediana no es una media). Decide usar el **bootstrap no paramétrico**.

- (1) Describe el procedimiento: ¿qué significa una “remuestra bootstrap” y cómo se construye? ¿Cuántas se generan y qué se calcula en cada una?
- (2) Explica el método del *percentil*: a partir de las B medianas bootstrap $\hat{m}_1^*, \dots, \hat{m}_B^*$, ¿cómo se obtiene el IC del 95 %?
- (3) ¿Por qué el bootstrap “funciona”? Enuncia el principio del *plug-in*: la distribución empírica $\hat{F}_n$ reemplaza a la desconocida F , y el remuestreo imita el muestreo de la población. Menciona una limitación.

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