Probabilidad y Estadística (C)

Segundo Parcial – Tema

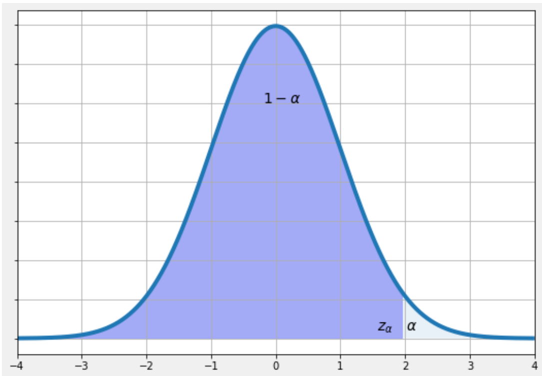
04 de julio de 2023

E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	Calificación

Nombre y Apellido: ______________________________

Libreta: ______________________________

- Tiene **cuatro horas** para realizar el examen.
- En los ejercicios E_1 a E_8, debe rodear/marcar con claridad la opción que considere correcta. Evitar redondeos en cuentas intermedias. Redondear al final y considerar 4 posiciones decimales. Estos ejercicios valen 10 puntos cada uno.
- El ejercicio E_9 debe resolverse en hoja aparte y vale 20 puntos.
- Para aprobar, se requiere un mínimo de 60 puntos.
- En todos los casos, los percentiles se indican a derecha, independientemente de la distribución. Por ejemplo, en el caso normal, z_α denota $\Phi(1-\alpha)$, como se ve en la imagen.



---

1. **Ejercicio E_1**:
La siguiente tabla muestra los resultados de 300 tiradas de un dado cargado.

Número	Frecuencia
1	10
2	60
3	130
4	53
5	17
6	30

Indicar el valor de la media podada al 17%. Recordar que

$$\bar{x}_\alpha = \frac{x_{([n\alpha]+1)} + \ldots + x_{(n-[n\alpha])}}{n-2[n\alpha]}$$

donde [ ] denota la parte entera.

a. 3.3233 **b.** 3.2697 **c.** 3.228 **d.** 3.1515

Respuesta
Buscamos el valor de la media podada al 17%. Es decir,

$$\begin{aligned}
\bar{x}_{17\%} &= \frac{x_{([300\cdot 17\%]+1)} + \ldots + x_{(300-[300\cdot 17\%])}}{300 - 2[300 \cdot 17\%]} \\
&= \frac{x_{(51+1)} + \ldots + x_{(300-51)}}{300 - 2 \cdot 51} \\
&= \frac{x_{(52)} + \ldots + x_{(249)}}{198} \\
&= \frac{2 + \ldots + 4}{198} \\
&= \frac{624}{198} \\
&= 3.1515
\end{aligned}$$

2. **Ejercicio E_2**:
Considerar el siguiente conjunto de observaciones $x_1, \ldots, x_{12}$

$$31, 65, 46, 50, 83, 78, 98, 25, 81, 94, 89, 68,$$

obtenidos a partir de una variable aleatoria X. Sea $y_i = ax_i + 7.54$ con a un número real positivo. Indicar el valor de a para que la mediana de $y_1, \ldots, y_{12}$ sea 843.39.

a. 18.99 **b.** 11.45 **c.** 12.88 **d.** 10.47

Respuesta
Las observaciones y_i son tales que $y_i = ax_i + 7.54$ para todo i. Como a es positivo, la mediana del conjunto de observaciones y_i cumple que

$$\tilde{y} = a\tilde{x} + 7.54$$

siendo $\tilde{x}$ la mediana de $x_1, \ldots, x_{12}$.

Busquemos $\tilde{x}$. Para ello, ordenamos los datos

$$31, 65, 46, 50, 83, 78, 98, 25, 81, 94, 89, 68,$$

y notemos que como 12 es par, la mediana resulta el promedio de las dos observaciones centrales de la muestra ordenada. Es decir, $\tilde{x} = 73$.

Finalmente,

$$a = \frac{\tilde{y} - 7.54}{\tilde{x}} = \frac{843.39 - 7.54}{73} = \frac{835.85}{73} = 11.45.$$

3. **Ejercicio E_3**:
Sea la muestra de tamaño 10

$$0.11, 1.92, 2.43, 5.09, 6.44, 7.98, 10.85, 13.23, 14.27, 14.3,$$

generada a partir de la variable aleatoria X con función de densidad

$$f_X(x, \theta) = \frac{3x^2}{512\theta^3} I_{[0,8\theta]}(x),$$

donde $\theta > 0$ es un parámetro desconocido. Indicar el valor de la estimación por máxima verosimilitud de θ a partir de los datos de la muestra.

a. 7.662 **b.** 14.3 **c.** 61.296 **d.** 1.7875

Respuesta
El estimador por máxima verosimilitud de 8θ es $\widehat{8\theta}_{MV} = \text{máx}_{1\leq i\leq n} \{X_i\}$. Luego, por equivarianza de los estimadores de máxima verosimilitud,

$$\widehat{\theta}_{MV} = \frac{\text{máx}_{1\leq i\leq n} \{X_i\}}{8}.$$

A partir de los datos, entonces, la estimación por máxima verosimilitud de θ es

$$\hat{\theta}_{MV_{obs}} = \frac{\text{máx}_{1\leq i\leq n} \{x_i\}}{8} = \frac{14.3}{8} = 1.7875.$$

4. **Ejercicio E_4**:
Sea p la proporción real de individuos de una población que consumen un producto A. Para estimar p se eligen n personas al azar de la población y se les pregunta si consumen o no el producto A. Se propone estimar p utilizando el método de los momentos (primer momento) y dicho estimador se nota como $\widehat{p}_n$. Suponiendo que $p = 0.22$, indicar el valor de a para que esta afirmación sea verdadera

$$\sqrt{n}\,(\hat{p}_n - 0.22) \xrightarrow{\mathcal{D}} N(0, a).$$

a. 0.1716 **b.** 0.22 **c.** 1 **d.** 0.0538

Respuesta
Sea

$$X_i = \begin{cases} 1 & \text{si la } i\text{−ésima persona encuestada dice consumir el producto A} \\ 0 & \text{en otro caso.} \end{cases}$$

para i entre 1 y n. Asumamos que las X_i son v.a.i.i.d. con distribución $Be(p)$. Luego, $E(X_i) = p$ y $V(X_i) = p(1-p)$ para todo i.

El estimador de p por el método de los momentos (primer momento) es

$$\widehat{p}_n = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}_n,$$

y su esperanza y varianza son

$$E(\widehat{p}_n) = E(\overline{X}_n) = p, \quad V(\widehat{p}_n) = V(\overline{X}_n) = \frac{p(1-p)}{n}.$$

Con $p = 0.22$, $E(\widehat{p}_n) = 0.22$ y $V(\widehat{p}_n) = \frac{0.1716}{n}$. Por el Teorema Central del límite, entonces

$$\sqrt{n}\frac{(\hat{p}_n - 0.22)}{\sqrt{0.1716}} \xrightarrow{\mathcal{D}} N(0, 1).$$

Luego,

$$\sqrt{n}\,(\hat{p}_n - 0.22) \xrightarrow{\mathcal{D}} N(0, 0.1716).$$

Por lo tanto $a = 0.1716$.

5. **Ejercicio E_5**:
La distribución del índice de colesterol en cierta población es $N(\mu, \sigma^2)$. Se hacen análisis a 10 personas elegidas al azar entre esta población y se obtienen los siguientes valores

$$2.24, 2.72, 2.64, 2.9, 2.69, 2.37, 2.45, 2.43, 2.48, 2.41.$$

Elegir cuál es el intervalo de confianza para μ de nivel 0.92. Se deja una lista de valores que podrían ser útiles para el problema:

$$z_{\alpha/2} = 1.7507, \quad z_\alpha = 1.4051, \quad t_{n-1,\alpha/2} = 1.9727, \quad t_{n,\alpha/2} = 1.9481, \quad \sum_{i=1}^{10}(x_i - \overline{x}_{10})^2 = 0.3536.$$

a. $(2.4094; 2.6566)$ **b.** $(2.4109; 2.6551)$ **c.** $(2.4449; 2.6211)$ **d.** $(2.4233; 2.6427)$

Respuesta
La distribución del índice de colesterol en cierta población es $N(\mu, \sigma^2)$ con σ^2 desconocido. Luego, el intervalo de confianza para μ de nivel $1-\alpha$ será de la forma

$$IC_\mu = \left(\overline{X}_n - \frac{S \cdot t_{n-1,\alpha/2}}{\sqrt{n}}; \overline{X}_n + \frac{S \cdot t_{n-1,\alpha/2}}{\sqrt{n}}\right)$$

siendo $S = \sqrt{\frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1}}$ y $t_{n-1,\alpha/2}$ el percentil $1-\alpha/2$ de una T con $n-1$ grados de libertad.

Con los datos del problema,

$$\overline{x}_{10} = 2.533, \quad s = 0.1982171, \quad \alpha = 0.08, \quad t_{10-1,0.08/2} = 1.9727$$

Luego

$$IC_\mu = \left(2.533 - \frac{0.1982171 \cdot 1.9727}{\sqrt{10}}; 2.533 + \frac{0.1982171 \cdot 1.9727}{\sqrt{10}}\right) = (2.4094; 2.6566).$$

6. **Ejercicio** E_6:
Una droga cura cierta enfermedad con probabilidad p. En una prueba con 60 enfermos, se curaron 45. Elegir el intervalo de confianza para p de nivel asintótico 0.91.

Se deja una lista de valores que podrían ser útiles para el problema:

$$z_\alpha = 1.3408, \quad z_{\alpha/2} = 1.6954, \quad t_{n-1,\alpha} = 1.3568, \quad -t_{n,\alpha} = 1.3566.$$

a. $(0.6742; 0.8258)$ **b.** Falta información, se necesita saber s. **c.** $(0.6552; 0.8448)$ **d.** $(0.675; 0.825)$

Respuesta
El intervalo de confianza para p de nivel asintótico $1-\alpha$ es de la forma

$$IC_p = \left(\overline{X}_n - \frac{\sqrt{\overline{X}_n \cdot (1-\overline{X}_n)} \cdot z_{\alpha/2}}{\sqrt{n}} ; \overline{X}_n + \frac{\sqrt{\overline{X}_n \cdot (1-\overline{X}_n)} \cdot z_{\alpha/2}}{\sqrt{n}} \right)$$

Con los datos del problema,

$$\overline{x}_{60} = 0.75, \quad \alpha = 0.09, \quad z_{0.09/2} = 1.6954$$

Luego

$$\begin{aligned} IC_p &= \left(0.75 - \frac{\sqrt{0.75 \cdot (1-0.75)} \cdot 1.6954}{\sqrt{60}} ; 0.75 + \frac{\sqrt{0.75 \cdot (1-0.75)} \cdot 1.6954}{\sqrt{60}} \right) \\ &= (0.6552; 0.8448). \end{aligned}$$

7. **Ejercicio** E_7:
Se sabe que la proporción de estudiantes de la FCEyN que recibe algún tipo de beca o estímulo económico es 0.4. Se sospecha que dicha proporción no es tan alta en el caso de los estudiantes de Computación, para lo que se realiza un test de hipótesis de nivel asintótico $\alpha = 0.09$. Indicar la probabilidad aproximada de no encontrar evidencia para rechazar $H_0 : p \geq 0.4$ en una muestra de tamaño 63, cuando en realidad la verdadera proporción de estudiantes de Computación que recibe algún tipo de beca o estímulo económico es 0.29.

Se deja una lista de valores que podrían ser útiles para el problema:

$$z_\alpha = 1.3408, \quad z_{\alpha/2} = 1.6954, \quad -z_\alpha = -1.3408, \quad -z_{\alpha/2} = -1.6954.$$

a. 0.6832 **b.** 0.2053 **c.** 0.3168 **d.** 0.09

Respuesta
Sea

$$X_i = \begin{cases} 1 & \text{si el } i\text{−ésimo estudiante de Comp. recibe algún tipo de beca o estímulo económico} \\ 0 & \text{en otro caso.} \end{cases}$$

para i entre 1 y n. Asumamos que las X_i son v.a.i.i.d. con distribución $Be(p)$, con $E(X_i) = p$ y $V(X_i) = p(1-p)$ para todo i.

El test de nivel asintótico 0.09 que se plantea es para las hipótesis

$$H_0 : p \geq 0.4 \quad \text{vs.} \quad H_1 : p < 0.4,$$

tiene estadístico $T = \sqrt{n}\frac{\overline{X}_n - 0.4}{\sqrt{0.4(1-0.4)}}$, cuya distribución asintótica es normal estándar y da lugar a la región de rechazo $(-\infty; -z_{0.09})$ con $-z_{0.09} = -1.3408$.

Si el verdadero p es 0.29, la probabilidad aproximada de no rechazar H_0 con $n = 63$ es

$$\begin{aligned}
P_{p=0.29}(T > -z_{0.09}) &= P_{p=0.29}\left(\sqrt{n}\frac{\overline{X}_n - 0.4}{\sqrt{0.4(1-0.4)}} > -z_{0.09}\right) \\
&= P_{p=0.29}\left(\overline{X}_n > \frac{-z_{0.09}\sqrt{0.4(1-0.4)}}{\sqrt{n}} + 0.4\right) \\
&= P_{p=0.29}\left(\sqrt{n}\frac{\overline{X}_n - 0.29}{\sqrt{0.29(1-0.29)}} > \frac{-z_{0.09}\sqrt{0.4(1-0.4)}}{\sqrt{0.29(1-0.29)}} + \frac{\sqrt{n}(0.4-0.29)}{\sqrt{0.29(1-0.29)}}\right) \\
&\approx 1 - \phi\left(\frac{-z_{0.09}\sqrt{0.4(1-0.4)}}{\sqrt{0.29(1-0.29)}} + \frac{\sqrt{n}(0.4-0.29)}{\sqrt{0.29(1-0.29)}}\right) \\
&\approx 1 - \phi\left(\frac{-1.3408\sqrt{0.4(1-0.4)}}{\sqrt{0.29(1-0.29)}} + \frac{\sqrt{63}(0.4-0.29)}{\sqrt{0.29(1-0.29)}}\right) \\
&\approx 0.3168.
\end{aligned}$$

8. **Ejercicio** E_8:
Sea el test para μ en una población $X \sim N(\mu, \sigma^2)$ de nivel 0.05 con hipótesis

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0$$

y considerar el estadístico y la región de rechazo que hemos estudiado para este problema cuando no se conoce la varianza poblacional. Se tiene una muestra de tamaño 14 que da lugar a un estadístico observado igual a 1.7422624. Indicar la única afirmación correcta a partir de estos datos.

Se deja una lista de valores que podrían ser útiles para el problema:

$$z_{0.05} = 1.6448536, \quad z_{0.025} = 1.959964, \quad -z_{0.05} = -1.6448536, \quad -z_{0.025} = -1.959964.$$

$$t_{14,0.05} = 1.7613101, \quad t_{14,0.025} = 2.1447867, \quad -t_{14,0.05} = -1.7613101, \quad -t_{14,0.025} = -2.1447867.$$

$$t_{13,0.05} = 1.7709334, \quad t_{13,0.025} = 2.1603687, \quad -t_{13,0.05} = -1.7709334, \quad -t_{13,0.025} = -2.1603687.$$

a. Hay evidencia suficiente para rechazar H0 al nivel dado.

b. NO hay un test de nivel 0.05 para esas hipótesis.

c. NO hay evidencia suficiente para rechazar H0 al nivel dado.

d. NO es posible decidir sobre el rechazo sin conocer s.

Respuesta
El test para μ en una población $X \sim N(\mu, \sigma^2)$ con σ^2 desconocido de nivel 0.05 con hipótesis

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0$$

tiene como estadístico a

$$T = \sqrt{n}\frac{\overline{X}_n - \mu_0}{S}$$

y cuya región de rechazo es $(t_{n-1,\alpha}, +\infty)$ siendo $t_{n-1,\alpha}$ el percentil que acumula área $1-\alpha$ en una distribución T con $n-1$ grados de libertad.

Según se indica, con una muestra de tamaño 14 se obtiene $T_{obs} = 1.7422624$. La región de rechazo para $\alpha = 0.05$ y $n = 14$ resulta $(1.7709334, +\infty)$ y, como el estadístico observado T_{obs} no pertenece a la región de rechazo, entonces la única respuesta correcta es “NO hay evidencia suficiente para rechazar H0 al nivel dado.”.

9. **Ejercicio** E_9:
Un grupo de estudiantes de Computación en FCEyN tiene la hipótesis de que ChatGPT NO es una buena herramienta (aún) para resolver problemas de matemática de nivel escolar porque, por lo general, se equivoca (incluso en cosas sencillas). El problema es, desde luego, muy amplio, así que decidieron restringirlo al caso de ejercicios típicos de polinomios de nivel escolar. Para ello, plantearon 40 ejercicios, todos de similar dificultad y en sesiones independientes, y ChatGPT respondió 34 incorrectamente.

- Indicar cuál es el parámetro de interés, la población y todos los supuestos que sean necesarios para tratar este problema.
- Dar una estimación puntual del parámetro de interés indicando qué estimador se usa y por qué.
- Dar una estimación del parámetro de interés por intervalo de confianza. Usar nivel $\alpha = 0.93$ (e indicar si el intervalo que se construye es de nivel exacto o asintótico).

- ¿Es posible usar el intervalo de confianza hallado para responder sobre algún test de hipótesis para el parámetro de interés? Si es posible, indicar a qué test se refiere y con qué nivel; si no es posible, explicar por qué.

Cada uno de estos ítemes vale 5 puntos y son calificados de forma independiente (i.e., no se penalizan errores sucesivas veces).

Respuesta
Sea

$$X_i = \begin{cases} 1 & \text{si la } i\text{–ésima pregunta(*) contestada por ChatGPT es respondida erróneamente} \\ 0 & \text{en otro caso.} \end{cases}$$

para i entre 1 y n. El (*) indica que por "pregunta" nos referimos específicamente a ejercicios típicos de polinomios de nivel escolar, todos de similar dificultad.

Para lo que sigue, debemos asumir que las X_i son v.a.i.i.d. con distribución $Be(p)$, siendo p la probabilidad de que ChatGPT responda incorrectamente una pregunta. El parámetro de interés es p.

El estimador de p por el método de los momentos (primer momento) es

$$\widehat{p}_n = \frac{\sum_{i=1}^n X_i}{n} = \overline{X}_n,$$

y una estimación de p a partir de él es $\overline{x}_{40} = 0.85$.

El intervalo de confianza para p de nivel asintótico $1 - \alpha$ es de la forma

$$IC_p = \left(\overline{X}_n - \frac{\sqrt{\overline{X}_n \cdot (1 - \overline{X}_n)} \cdot z_{\alpha/2}}{\sqrt{n}} ; \overline{X}_n + \frac{\sqrt{\overline{X}_n \cdot (1 - \overline{X}_n)} \cdot z_{\alpha/2}}{\sqrt{n}} \right)$$

Con los datos del problema,

$$\overline{x}_{40} = 0.85, \quad \alpha = 0.07, \quad z_{0.07/2} = 1.8119$$

Luego

$$\begin{aligned} IC_p &= \left(0.85 - \frac{\sqrt{0.85 \cdot (1 - 0.85)} \cdot 1.8119}{\sqrt{40}} ; 0.85 + \frac{\sqrt{0.85 \cdot (1 - 0.85)} \cdot 1.8119}{\sqrt{40}} \right) \\ &= (0.7477; 0.9523). \end{aligned}$$

Sí, es posible usar el intervalo hallado para responder acerca del rechazo en un test de hipótesis bilateral para p de nivel asintótico 0.07. Por ejemplo, si se testea $p = p_0$ contra $p \neq p_0$, basta ver si p_0 pertenece (o no) al intervalo hallado para concluir, con esa muestra y a nivel asintótico 0.07, que no se rechaza H_0 (o sí).