Probabilidad y Estadística (C)

Segundo Parcial – Tema 10

04 de julio de 2023

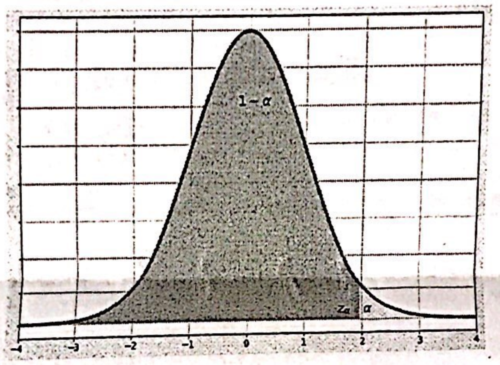
¡Muy bien!

E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	Calificación
✓	✓	✓	✓	✓	✓	✓	✓	✓	100

NOMBRE Y APELLIDO: Pedro Fuentes Urfeig

LIBRETA: [redacted]

- 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 $\Phi^{-1}(1-\alpha)$, como se ve en la imagen.



---

1. **Ejercicio E_1:**
La siguiente tabla muestra los resultados de 600 tiradas de un dado cargado.

Número	Frecuencia
1	239
2	142
3	29
4	35
5	101
6	54

Indicar el valor de la media podada al 23%. 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. 2.4048 **b.** 2.5568 **c.** 2.6317 **d.** 2.1512

---

2. **Ejercicio E_2:**
Considerar el siguiente conjunto de observaciones $x_1, \cdots, x_{14}$

$$72, 85, 86, 69, 44, 74, 82, 96, 21, 40, 51, 42, 94, 68,$$

obtenidos a partir de una variable aleatoria X. Sea $y_i = ax_i + 6.11$ con a un número real positivo. Indicar el valor de a para que la mediana de $y_1, \cdots, y_{14}$ sea 593.375.

a. 9.22

b. 8.33

c. 7.75

d. 14.44

---

3. **Ejercicio E_3:**
Sea la muestra de tamaño 11

$$0.56, 0.7, 1.82, 1.91, 3.06, 3.34, 8.6, 11.02, 11.4, 15.05, 15.75,$$

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

$$f_X(x,\theta) = \frac{3x^2}{216\theta^3} I_{[0,6\theta]}(x),$$

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. 15.75

b. 2.625

c. 39.9327

d. 6.6555

---

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. 0.22

b. 1

c. 0.0538

d. 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

$$2.42, 2.4, 2.61, 2.61, 2.64, 2.67, 2.47, 2.65, 2.69, 2.36.$$

Elegir cuál es el intervalo de confianza para μ de nivel 0.9. Se deja una lista de valores que podrían ser útiles para el problema:

$$z_{\alpha/2} = 1.6449, \quad z_\alpha = 1.2816, \quad t_{n-1,\alpha/2} = 1.8331, \quad t_{n,\alpha/2} = 1.8125, \quad \sum_{i=1}^{10}(x_i - \overline{x}_{10})^2 = 0.1412.$$

a. (2.4802; 2.6238)

b. (2.5012; 2.6028)

c. (2.4794; 2.6246)

d. (2.4869; 2.6171)

---

6. **Ejercicio E_6:**
Una droga cura cierta enfermedad con probabilidad p. En una prueba con 64 enfermos, se curaron 48. Elegir el intervalo de confianza para p de nivel asintótico 0.96.

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

$$z_\alpha = 1.7507, \quad z_{\alpha/2} = 2.0537, \quad t_{n-1,\alpha} = 1.7794, \quad -t_{n,\alpha} = 1.7789.$$

a. (0.6552; 0.8448)

b. (0.6537; 0.8463)

c. Falta información, se necesita saber s.

d. (0.6388; 0.8612)

---

7. **Ejercicio E_7:**
Se sabe que la proporción de estudiantes de la FCEyN que recibe algún tipo de beca o estímulo económico es 0.8. 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 $\alpha = 0.07$. Indicar la probabilidad aproximada de no encontrar evidencia para rechazar $H_0 : p \geq 0.8$ en una muestra de tamaño 91, cuando en realidad la verdadera proporción de estudiantes de Computación que recibe algún tipo de beca o estímulo económico es 0.76.

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

$$z_\alpha = 1.4758, \quad z_{\alpha/2} = 1.8119, \quad -z_\alpha = -1.4758, \quad -z_{\alpha/2} = -1.8119.$$

a. 0.0374 b. 0.6875 c. 0.3125 d. 0.07

---

8. **Ejercicio E_8:**
Sea el test para μ en una población $X \sim N(\mu, \sigma^2)$ de nivel 0.02 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 13 que da lugar a un estadístico observado igual a 2.5676928. Indicar la única afirmación correcta a partir de estos datos.

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

$$z_{0.02} = 2.0537489, \quad z_{0.01} = 2.3263479, \quad -z_{0.02} = -2.0537489, \quad -z_{0.01} = -2.3263479.$$

$$t_{13,0.02} = 2.2816036, \quad t_{13,0.01} = 2.6503088, \quad -t_{13,0.02} = -2.2816036, \quad -t_{13,0.01} = -2.6503088.$$

$$t_{12,0.02} = 2.3027217, \quad t_{12,0.01} = 2.680998, \quad -t_{12,0.02} = -2.3027217, \quad -t_{12,0.01} = -2.680998.$$

a. NO hay un test de nivel 0.02 para esas hipótesis.

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

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

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

---

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 50 ejercicios, todos de similar dificultad y en sesiones independientes, y ChatGPT respondió 29 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 $\alpha = 0.94$ (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).

Pedro Fuentes Urfeig [redacted] Hoja 7.

9) i) Sea $X_i = \begin{cases} 1 \text{ si CHATGPT resolvió incorrectamente} \\ 0 \text{ si no} \end{cases} \sim Be(p)$. ($X_i$ son VA. IID)

p en este caso representa ~~el~~ por la proporción de ejercicios incorrectos que resolvió ~~del~~ CHATGPT. P es mi parámetro de interés.

Tengo una muestra aleatoria de X, $X_1 \dots X_{50}$ tal que $\sum_{i=1}^{50} X_i = 29$.

$\frac{1}{50}\sum_{i=1}^{50} X_i = 0{,}58$.

ii) Busco un estimador de p con el estimador del primer momento y lo llamo $\hat{p}_n$

$E[X_i] = \overline{X}$

$\hookrightarrow p = \overline{X}$

$\hat{p}_n = \overline{X}$. Para esta muestra, uso $\hat{p}_n = 0{,}58$.

.) Sé que este estimador es consistente e insesgado.

Consistente

$\overline{X} \to E[X]$ por LGN

$\hookrightarrow \hat{p}_n \to p$

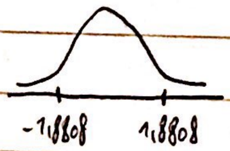
Insesgado

$E[\hat{p}_n] = E(\overline{X}) = E[X] = p$.

iii) Busco un intervalo de confianza para p de nivel $\alpha = 0{,}06$. En este caso, el intervalo es asintótico.

Sea $T = \frac{\hat{p}_n - p}{\sqrt{\hat{p}_n - \hat{p}_n^2}}\sqrt{n} \overset{(a)}{\approx} N(0,1)$ (por TCL) mi pivote.

.) $Z_{\frac{\alpha}{2}} = Z_{0{,}03} = 1{,}8808$



$\Rightarrow P\left(-1{,}8808 \le \frac{\hat{p}_n - p}{\sqrt{\hat{p}_n - \hat{p}_n^2}}\sqrt{n} \le 1{,}8808\right) \simeq 0{,}94$

Luego, mi intervalo de confianza de nivel asintótico 0,94 es el siguiente (con los datos de mi muestra).

$\left[0{,}58 - 1{,}8808 \cdot \frac{\sqrt{0{,}58 \cdot 0{,}42}}{\sqrt{50}} ; 0{,}58 + 1{,}8808 \cdot \frac{\sqrt{0{,}58 \cdot 0{,}42}}{\sqrt{50}}\right] \Rightarrow \left[0{,}4487 ; 0{,}7113\right]$

iv). Es posible construir un Test de hipótesis para p de nivel asintótico $\alpha = 0.06$.

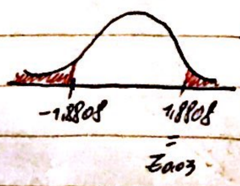
Hipótesis

$H_0: p = 0{,}58$ vs $H_1: p \neq 0{,}58$.

Estadístico

$$T = \frac{\hat{p}_n - 0{,}58}{\sqrt{0{,}58 \cdot 0{,}42}} \sqrt{n} \underset{\text{Bajo } H_0}{\overset{(a)}{\sim}} N(0,1) \text{ por TCL.}$$

Región de Rechazo



Area (■) = $\alpha = 0{,}06$.

Rechazo H_0 si $|T| > 1{,}8808$.

·) Luego, veo que si evalúo mi estadístico con $\hat{p}_n = 0{,}4487$ y $\hat{p}_n = 0{,}7113$ (los bordes de mi intervalo de confianza) obtengo $-1{,}8810$ y $1{,}8811$, respectivamente, lo cual es bastante similar a los ~~los~~ "bordes" de mi región de rechazo.

·) Luego, relaciono mi intervalo de confianza con el Test de hipótesis ya que el intervalo de confianza del parámetro p coincide con la región de No rechazo del Test de hipótesis de nivel $\alpha = 0{,}06$ (el mismo nivel que el IdC).