1	2	3	4	Calificación
30	18	27	14	89

A

TEMA 2

Probabilidad y Estadística (C)

Segundo parcial - 30/06/2022

Complete esta hoja y entréguela con el resto del examen. ***Realizar cada ejercicio en hoja separada.*** *Escribir el nombre en cada una. Al retirarse debe firmar una hoja de asistencia.*

APELLIDO Y NOMBRES: Pagés Julieta Nº DE LIBRETA: 1691/21

mail: juli.b.pages@gmail.com FIRMA:

Turno: Tarde: 14 a 17 hs | Noche: 19 a 22 hs   Nº de hojas entregadas (sin enunciado):

Criterio de aprobación: Para aprobar este examen es necesario sumar al menos 60 puntos y tener un ejercicio bien resuelto.

Definir con palabras los eventos y/o las variables aleatorias involucradas, nombres y parámetros de las distribuciones. Resolver usando al menos 4 decimales. ***Justifique claramente sus afirmaciones.***

---

1. (30 puntos) Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución $U[0, \beta]$. Consideremos tres estimadores de β:

   - $\widehat{\beta}_1$ el estimador de momentos,
   - $\widehat{\beta}_2$ el estimador de máxima verosimilitud y
   - $\widehat{\beta}_3 = k\widehat{\beta}_2$ un estimador insesgado basado en $\widehat{\beta}_2$, siendo $k \in \mathbb{R}$.

   (a) (17 puntos) Hallar el sesgo de los tres estimadores. ¿Cuáles son insesgados?

   (b) (8 puntos) Verificar que $V(\widehat{\beta}_2) = \dfrac{n}{(n+2)(n+1)^2}\beta^2$.

   (c) (5 puntos) Probar que $ECM(\widehat{\beta}_3) \leq ECM(\widehat{\beta}_2) \leq ECM(\widehat{\beta}_1)$ para todo $n \in \mathbb{N}$. ¿Estos estimadores son consistentes? Justificar.

2. (a) (15 puntos) Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $\mathcal{P}(\lambda)$. Hallar un intervalo de confianza para λ, basado en el estimador de momentos, de nivel asintótico $1 - \alpha$.

   (b) (5 puntos) El número de llamadas que recibe por día una central telefónica son variables aleatorias independientes con distribución de Poisson de parámetro λ. Se registró la cantidad de llamadas diarias durante 30 días, obteniendose las siguientes observaciones:

35	41	38	40	34	36	41	48	42	46	41	35	37	38	42
39	37	41	35	37	38	42	43	44	60	41	38	40	34	36

   Utilizando el ítem (a), hallar un intervalo de confianza de nivel asintótico 0.95 para λ, basado en esta muestra.

   Obs: Si lo necesita puede usar que $\sum_{i=1}^{30} x_i = 1199$.

3. (35 puntos) Se sabe que la cantidad de colesterol de los hombres de entre 40 y 50 años está dada por una variable aleatoria con distribución normal. En una localidad de Jujuy esta variable es $N(204, 81)$. Se quiere estudiar si en una localidad de Tucumán la media de colesterol difiere de la media de la localidad jujeña. Para ello, vamos a construir un test de hipótesis basado en una muestra de 15 hombres (de Tucumán) que permita tomar esa decisión, con un nivel de significación de 0.05.

(a) (13 puntos) Supongamos que el desvío estándar en ambos pueblos es el mismo. Definir las variables aleatorias involucradas, las hipótesis del test, el estadístico y la regla de decisión. ¿Para qué valores del promedio muestral $\bar{x}_{15}$ se puede concluir que las medias de colesterol de los hombres de ambas localidades difieren a nivel $\alpha = 0.05$?

(b) (10 puntos) A partir del test hallado en el ítem anterior, calcular la probabilidad de tomar una decisión incorrecta si la verdadera media fuera $\mu = 201$.

(c) (5 puntos) ¿Qué se concluye en base a la muestra observada en la localidad de Tucumán?

198 185 214 189 206 212 193 209 196 211 189 198 211 191 199

Obs: Si lo necesita puede usar que $\sum_{i=1}^{15} x_i = 3001$.

(d) (7 puntos) Suponemos que el desvío estándar de ambas localidades ya no es el mismo. Utilizar los datos del ítem anterior para obtener un intervalo de confianza del 90% para la media de colesterol de los hombres de Tucumán.

4. (a) (8 puntos) Considerar los siguientes 130 datos obtenidos de una encuesta que informa la cantidad de celulares por cada hogar visitado. Calcular la mediana y la media α podada al 10% de la cantidad de celulares por hogar.

# celulares	# hogares
0	23
1	51
2	31
3	17
4	8

(b) (7 puntos) Se tiene una muestra de 30 datos $x_1, \ldots, x_{30}$ cuya desviación estándar muestral es 7 y se tiene otra muestra de 20 datos $y_1, \ldots, y_{20}$ con desviación estándar muestral igual a 9. Se forma una muestra de 50 datos uniendo las dos anteriores $x_1, \ldots, x_{30}, y_1, \ldots, y_{20}$. Si se sabe que $\bar{x}_{30} = \bar{y}_{20}$, hallar la desviación estándar muestral de esta nueva muestra.

Recordar: llamamos varianza muestral de $z_1, \ldots, z_n$ a $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{z}_n)^2$.

Julieta Pagés   tema 2

① $X_1, \dots, X_n$ m.a con distribución $\mathcal{U}(0,\beta)$

a) $E(X_i) = \frac{0+\beta}{2}$ $\Rightarrow$ quiero el $\hat{\beta}_1$ (estimador de momentos) que cumpla que $\frac{\hat{\beta}_1}{2} = \bar{X} \leftrightarrow \boxed{\hat{\beta}_1 = 2\bar{X}}$ ✓

$E_\beta(\hat{\beta}_1) = E_\beta(2\bar{X}) = \frac{2\beta}{2} = \beta$ ✓

Por lo tanto $\hat{\beta}_1$ es insesgado, es decir el sesgo es 0 $(\underbrace{E_\beta(\hat{\beta}_1) - \beta}_{\text{sesgo}} = \beta - \beta = 0)$ ✓

Estimador de máxima verosimilitud:

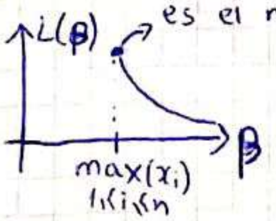
$\cancel{L(\theta) = \prod \frac{1}{\beta} = \frac{1}{\beta^n}}$ $L(\theta) = \prod_{i=1}^{n} \frac{1}{\beta} I_{(0,\beta)}(x_i) = \frac{1}{\beta^n} \prod_{i=1}^{n} I_{(0,\beta)}(x_i)$

$L(\beta) = \begin{cases} 1/\beta^n & \text{si } \overset{\text{es seguro}}{0 \leq x_i \leq \beta} \ (1 \leq i \leq n) \overset{*}{\Longleftrightarrow} \max_{1\leq i \leq n}(x_i) \leq \beta \\ 0 & \text{en otro caso} \end{cases}$

* $0 \leq x_i$ $(1 \leq i \leq n)$ es seguro que pasa porque al tomar la muestra, x_i va a ser un número que pertenece al intervalo de la uniforme, ya que esa es su distribución ✓

ahora $x_i \leq \beta$, $1 \leq i \leq n$ es lo mismo que decir $\max_{1\leq i\leq n}(x_i) \leq \beta$

Entonces ✓



Por lo tanto $\hat{\beta}_2 = \max_{1\leq i\leq n}(X_i)$ X_i V.A.

Busco su distribución:

$P(\max_{1\leq i\leq n}(X_i) \leq \tau) = (F_X(\tau))^n$

✓ $\Rightarrow f_{\max_{1\leq i\leq n}(X_i)}(\tau) = n(F_X(\tau))^{n-1} f_X(\tau) = n\left(\frac{\tau}{\beta}\right)^{n-1} \frac{1}{\beta} I_{(0,\beta)}(\tau) = \frac{n\,\tau^{n-1}}{\beta^n} I_{(0,\beta)}(\tau)$

$\Rightarrow E_\beta(\hat{\beta}_2) = E_\beta(\max_{1\leq i\leq n}(X_i)) = \int_0^\beta \tau \frac{n\,\tau^{n-1}}{\beta^n} d\tau = \frac{n}{(n+1)} \frac{\tau^{n+1}}{\beta^n}\Big|_0^\beta = \frac{n}{(n+1)} \beta$ ✓

Entonces el sesgo de $\hat{\beta}_2$ es $E_\beta(\hat{\beta}_2) - \beta = \boxed{\frac{n}{n+1}\beta - \beta}$ ✓ es decir NO es insesgado

$= \beta\left(\frac{-1}{n+1}\right)$

Para $\hat{\beta}_3$ tomo $K = \frac{n+1}{n}$, o sea $\hat{\beta}_3 = \frac{n+1}{n}\hat{\beta}_2 = \frac{n+1}{n}\max_{1\leq i\leq n}(x_i)$

Como vemos que $E\left(\frac{n+1}{n}\hat{\beta}_2\right) = \frac{\cancel{n+1}}{\cancel{n}}\frac{\cancel{n}}{\cancel{n+1}}\beta = \beta = E(\hat{\beta}_3)$

o sea que el sesgo es $E(\hat{\beta}_3) - \beta = \beta - \beta = 0$, es decir que es insesgado. ✓

☒

b) $V(\hat{\beta}_2) = V(\max_{1\leq i\leq n}(x_i)) = E(\hat{\beta}_2^2) - (E(\hat{\beta}_2))^2 = \frac{n}{n+2}\beta^2 - \left(\frac{n}{n+1}\beta\right)^2 =$

$$E(Y^2) = \int_0^\beta \tau^2 \frac{n\tau^{n-1}}{\beta^n}d\tau = \int_0^\beta \frac{n\tau^{n+1}}{\beta^n}d\tau = \frac{n}{(n+2)}\frac{\tau^{n+2}}{\beta^n}\Big|_0^\beta = \frac{n}{n+2}\beta^2$$

$$= \beta^2\left(\frac{n}{n+2} - \frac{n^2}{(n-1)^2}\right) = \beta^2\left(\frac{(n-1)^2 n - (n+2)n^2}{(n+2)(n+1)^2}\right) = \beta^2\left(\frac{\cancel{n^3} - \cancel{n^3} - 2n^2 + 2n^2 + n}{(n+2)(n+1)^2}\right)$$

$$= \beta^2\left(\frac{n}{(n+2)(n+1)^2}\right)$$ como se quería probar ✓

☒

c) $ECM(\hat{\beta}_3) = V(\hat{\beta}_3) + (\underbrace{E_\beta(\hat{\beta}_3) - \beta}_{=0})^2 = V(\hat{\beta}_3) = E(\hat{\beta}_3^2) - (E(\hat{\beta}_3))^2$

~~Primero necesito su distribución.~~

~~$P\left(\frac{n+1}{n}\max_{1\leq i\leq n}(x_i) \leq y\right) = P\left(\max_{1\leq i\leq n}(x_i) \leq \frac{n}{n+1}y\right) = F_{\hat{\beta}_2}\left(\frac{n}{n+1}y\right)$~~

~~$\Rightarrow f_{\hat{\beta}_3}(y) = n F_x\left(\frac{n}{n+1}y\right)^{n-1} f_x\left(\frac{n}{n+1}y\right) = n\left(\frac{n}{(n+1)}\frac{y}{\beta}\right)^{n-1}\frac{1}{\beta}$~~

~~$F_{\hat{\beta}_3}(y) = F_{\hat{\beta}_2}\left(\frac{n}{n+1}y\right) = n$~~

Salía por propiedad de la Var (β^2)

$$= E\left(\left(\frac{n+1}{n}\max_{1\leq i\leq n}(x_i)\right)^2\right) - \beta^2 = E\left(\left(\frac{n+1}{n}\right)^2\max_{1\leq i\leq n}(x_i)^2\right) - \beta^2$$

$$\overset{\text{linealidad}}{=} \left(\frac{n+1}{n}\right)^2 E(\underbrace{\max_{1\leq i\leq n}(x_i)^2}_{=\hat{\beta}_2}) - \beta^2 = \left(\frac{n+1}{n}\right)^2\left(\frac{n}{n+2}\beta^2\right) - \beta^2$$

$$= \beta^2\left(\frac{(n+1)^2 n}{n^2(n+2)} - 1\right) = \beta^2\left(\frac{(n+1)^2}{n(n+2)} - 1\right)$$

$$= \beta^2\left(\frac{n^2 + 2n + 1 - n^2 - 2n}{n(n+2)}\right) = \boxed{\beta^2\left(\frac{1}{n(n+2)}\right) = ECM(\hat{\beta}_3)}$$

$$ECM(\hat{\beta}_3) = \beta^2\left(\frac{1}{n(n+2)}\right)$$

$$ECM(\hat{\beta}_2) = V(\hat{\beta}_2) + \left(E_\beta(\hat{\beta}_2) - \beta\right)^2 = \frac{n}{(n+2)(n+1)^2}\beta^2 + \beta^2\frac{1}{(n+1)^2}$$

$$= \beta^2\left(\frac{n}{(n+2)(n+1)^2} + \frac{1}{(n+1)^2}\right) = \beta^2\left(\frac{n+n+2}{(n+2)(n+1)^2}\right) = \beta^2\left(\frac{2n+2}{(n+2)(n+1)^2}\right)$$

$$\boxed{\cancel{= \frac{3n\beta}{(n+2)(n+1)^2}}} \quad = \beta^2\left(\frac{2(n+1)}{(n+2)(n+1)^2}\right) = \beta^2\left(\frac{2}{(n+2)(n+1)}\right)$$

$$ECM(\hat{\beta}_1) = V(\hat{\beta}_1) + \underbrace{\left(E_\beta(\hat{\beta}_1) - \beta\right)^2}_{=0} = V(2\bar{X}) = 4V(\bar{X}) = \frac{4}{n}\frac{(\beta)^2}{12} = \frac{\beta^2}{3n}$$

Ahora:

$ECM(\hat{\beta}_2) \leq ECM(\hat{\beta}_1) \Leftrightarrow$ $\cancel{\frac{3n\beta^2}{(n+2)(n+1)^2} < \frac{\beta^2}{3n} \Leftrightarrow \frac{1}{\beta(n+2)(n+1)^2} < 1}$ $(\beta \neq 0)$

(*) (lo hago atrás)

ESTO es verdadero para todo n

Por otro lado:

puedo tachar lo porque $\beta \neq 0$

$$ECM(\hat{\beta}_3) \leq ECM(\hat{\beta}_2) \Leftrightarrow \cancel{\beta^2}\left(\frac{1}{n(n+2)}\right) \leq \cancel{\beta^2}\left(\frac{2}{(n+2)(n+1)}\right)$$

$$\Leftrightarrow 1 \leq \frac{2n}{n+1}$$ para $n=1$ $\frac{2n}{n+1} = \frac{2}{2} = 1 \geq 1$

y desde ahí n solo puede ser mayor, y $\frac{2n}{n+1}$, por lo tanto también va a ser cada vez mayor que 1. o sea, cumple $\forall n \in \mathbb{N}$

$$ECM(\hat{\beta}_1) = \frac{\beta^2}{3n} \xrightarrow{n \to \infty} 0$$

$$ECM(\hat{\beta}_2) = \beta^2\left(\frac{2}{(n+2)(n+1)}\right) \xrightarrow{n \to \infty} 0$$

$$ECM(\hat{\beta}_3) = \beta^2\left(\frac{1}{n(n+2)}\right) \xrightarrow{n \to \infty} 0$$

Por lo tanto, los tres estimadores son consistentes ✓

b)

② $X_1, \dots, X_n$ m.a $X_i \sim P(\lambda)$ $E(x) = V(x) = \lambda$

$\hat{\lambda}_{MO} = \overline{X}$ (visto en teórica)

¿Por qué?

$$\sqrt{n}\,\frac{\overline{X}-\lambda}{\sqrt{\lambda}} \overset{(a)}{\sim} N(0,1) \underset{\text{Por Slutsky}}{\Longrightarrow} \sqrt{n}\,\frac{\overline{X}-\lambda}{\sqrt{\overline{X}}} \overset{(a)}{\sim} N(0,1)$$

$$1-\alpha \underset{\text{aprox}}{\simeq} P\left(-z_{\alpha/2} \leq \sqrt{n}\,\frac{\overline{X}-\lambda}{\sqrt{\overline{X}}} \leq z_{\alpha/2}\right) = P\left(-z_{\alpha/2}\frac{\sqrt{\overline{X}}}{\sqrt{n}} + \overline{X} \leq \lambda \leq z_{\alpha/2}\frac{\sqrt{\overline{X}}}{\sqrt{n}} + \overline{X}\right)$$

Entonces, un intervalo de nivel asintótico $1-\alpha$ para λ es:

$$IC(\lambda) = \left[-z_{\alpha/2}\frac{\sqrt{\overline{X}}}{\sqrt{n}} + \overline{X} \,,\, z_{\alpha/2}\frac{\sqrt{\overline{X}}}{\sqrt{n}} + \overline{X}\right]$$

NOTA

b) Con los datos $\alpha = 0.05$, $n = 30$, $\bar{x} = \frac{1199}{30} = 39{,}967$

el intervalo de confianza de nivel asintótico 0,95 para λ es

$$IC_{obs}(\lambda) = \left[-z_{0.025} \frac{\sqrt{39{,}967}}{\sqrt{30}} + 39{,}967 \, , \, z_{0.025} \frac{\sqrt{39{,}967}}{\sqrt{30}} + 39{,}967 \right]$$

$$(z_{0.025} = 1{,}96) \qquad = [37{,}705 \, , \, 42{,}229]$$ ✓

Sería con 4 decimales, pero está bien.

27 pt

tema 2

a) ③ X_i = Cantidad de colesterol del i-ésimo hombre de entre
12pt 40 y 50 años de Jujuy ✓

$X_i \sim N(204, 81)$

Y_i = Cantidad de colesterol del i-ésimo hombre de entre 40 y 50 años de Tucumán

$Y_i \sim N(\mu, 81)$ ✓

$H_0: \mu = 204$ vs. $H_1: \mu \neq 204$ ✓

guarda, que tus V.A de la muestra son Y_i (tucumán)

Como Y_i es normal, el estadístico del test es:

$$T = \left| \sqrt{n}\, \frac{\bar{X} - \mu_0}{\sigma} \right| = \left| \sqrt{n}\, \frac{\bar{X} - 204}{9} \right|$$

y mi regla de desición:

$$RD \begin{cases} \text{Rechazo } H_0 & \left|\sqrt{n}\, \frac{\bar{X}-204}{9}\right| > z_{\alpha/2} = 1{,}96 \text{ (para } \alpha = 0.05) \\ \underline{\text{No}} \text{ rechazo } H_0 \text{ en caso contrario} \end{cases}$$ ✓

Entonces, para $n = 15$ y $\alpha = 0.05$

$$\left|\sqrt{15}\, \frac{\bar{X}-204}{9}\right| > 1{,}96 \iff \sqrt{15}\, \frac{\bar{X}-204}{9} > 1{,}96 \iff \bar{X} > \frac{9}{\sqrt{15}} \cdot 1{,}96 + 204 = 208{,}555$$

ó

$$-1{,}96 > \sqrt{15}\, \frac{\bar{X}-204}{9} \iff -1{,}96 \cdot \frac{9}{\sqrt{15}} + 204 > \bar{X} \iff \bar{X} < 199{,}445$$

Es decir que para $\bar{X}_{15} < 199{,}445$ ó $\bar{X}_{15} > 208{,}555$, puedo rechazar H_0, y por lo tanto concluir que las medias son distintas. ✓

con evidencia significativa a nivel 5% !

b) $P_{\mu=201}\left(\left|\sqrt{15}\ \frac{\bar{X}-204}{9}\right| < 1,96\right) = P_{\mu=201}\left(-1,96 < \sqrt{15}\ \frac{\bar{X}-201-3}{9} < 1,96\right)$

10 pt

$= P_{\mu=201}\left(-1,96 + \frac{\sqrt{15}.3}{9} < \sqrt{15}\ \frac{\bar{X}-201}{9} < 1,96 + \frac{\sqrt{15}.3}{9}\right) = \phi\left(1,96 + \frac{\sqrt{15}.3}{9}\right) - \phi\left(-1,96 + \frac{3\sqrt{15}}{9}\right)$

$= \phi(3,25) - \phi(-0,67) = \phi(3,25) - 1 + \phi(0,67) = 0,9994 - 1 + 0,7486$

$\boxed{= 0,748}$

La probabilidad de tomar la decisión incorrecta cuando $\mu=201$ ~~(FESHA~~ (No rechazar H_0 cuando no valía, el error de tipo II) es 0,748.

c) Con los datos de una muestra observada siendo

5 pt

$n=15 \qquad \bar{x} = \frac{3001}{15} = 200,067$

$T_{obs} = \left|\sqrt{15}\ \frac{200,067-204}{9}\right| = 1,6925 \not> 1,96$

Por lo tanto con los datos, NO puedo tomar la decisión de rechazar H_0, o sea, no puedo concluir que la media de la cantidad de colesterol de los hombres de entre 40 y 50 años difiere. a nivel 5%.

d) $\emptyset$

Julieta Pagés | Tema 2

④ a) X = cantidad de celulares en el hogar$_i$ $R_{X_i} = \{0,1,2,3,4\}$

~~$P(X_i \leq M) = 1/2 = 0.5$~~ La mediana es el valor que me deja la mitad de la muestra a un lado, y la otra mitad al otro, en este caso sería el valor que ~~ocupe su posición~~ esté entre las posiciones 65 y 66 o sea, es el $\boxed{1}$ ✓

Para la media α podada 10%, tengo que quitar el 10% de cada punta y calcular la media, en este caso $\frac{10}{100} \cdot 130 = 13$

o sea que la tabla quedaría:

cel	# hogares
0	10
1	51
2	31
3	12
4	0

✓

y la media sería: (el total ahora es 104)

$$\underbrace{0.\frac{10}{104}}_{P_x(0)\ \text{LaPlace}} + 1.\frac{51}{104} + 2.\frac{31}{104} + 3.\frac{12}{104} + 4.\frac{0}{104} = \boxed{1,433}$$ ✓

b) $x_1, \ldots, x_{30}$ muestra. $S_x = 7 \Rightarrow S_x^2 = 49$ $(\bar{x}_{30} = \bar{y}_{20})$

$y_1, \ldots, y_{20}$ muestra $S_y = 9 \Rightarrow S_y^2 = 81$

sea S el desvío estandar muestral que resulta de unir las dos muestras anteriores

$$S^2 = \frac{1}{49}\left(\underbrace{\sum_{i=1}^{30}(x_i - \bar{x}_{30})^2 + \sum_{i=1}^{20}(y_i - \bar{y}_{20})^2}\right)$$

puedo hacerlo ya que $\bar{x}_{30} = \bar{y}_{20}$, o sea que solo separé la gran sumatoria de 50 términos en dos, para separar lo que vino de X de lo que vino de Y

$$= \frac{30}{49}\frac{1}{30}\underbrace{\left(\sum_{i=1}^{30}(x_i - \bar{x}_{30})^2\right)}_{S_x^2} + \frac{20}{49}\frac{1}{20}\underbrace{\left(\sum_{i=1}^{20}(y_i - \bar{y}_{20})^2\right)}_{S_y^2}$$

29 19

$$= \frac{30}{49}\, 7^2 + \frac{20}{49}\, 9^2 = 63{,}061$$

Por lo tanto $\boxed{S = \sqrt{63{,}061} = 7{,}94}$ $\sim\checkmark$