13-2-17

Universidad Tecnológica Nacional Facultad Regional Buenos Aires Examen Final de Probabilidad y Estadística -15 de febrero de 2017

ApellidoyNombre : *Legajo :*

Ej. 1	Ej. 2	Ej. 3	Ej. 4	Teórico 1	Teórico 2	Nota

La condición mínima de aprobación es de tres puntos correctos. Todas las respuestas deberán estar debidamente justificadas.

Ejercicio 1 *Una reconocida marca de pilas lanza al mercado una nueva pila afirmando que la duración de estas es mayor a 120 hs. Una muestra de 23 pilas,arrojó un promedio de duración de 122 horas y un desvío estándar de 4 horas. Suponiendo que la duración tiene una distribución Normal.* 4

1. *Plantear las hipótesis del test y aplicarlo con nivel $\alpha = 0,05$.*
2. *Si el desvío estándar poblacional es conocido e igual a 4,5 con nivel de significación $\alpha = 0,05$, hallar el tamaño muestral tal que la probabilidad de equivocarse al tomar la decisión, cuando la verdadera media es de 120,5 horas, sea menor a 0,02.*

E(X^4)=7289

Ejercicio 2 *Sea X una variable aleatoria continua que verifica: $E(X) = 8$, $Var(X) = 21$, $E(X^4) = 7289$*
Sea $Y = X^2$ una variable aleatoria que representa "peso de un puma de cierta región de América del Sur"

1. *Calcular la esperanza y la varianza de Y.*
2. *Si se eligen al azar 80 pumas para ser transportados. ¿cuál es la probabilidad de sobrecarga sabiendo que la bodega tolera 6900 kg?*
3. *¿Cuántos pumas hay que elegir en forma independiente para tener al menos un 94% de probabilidad de que el peso promedio esté entre ~~79~~ 94 y ~~81~~ 96 kg?*

Ejercicio 3 *Los televisores de plasma tienen una cierta vida útil, debido a la pérdida del gas que compone la pantalla. Suponiendo que el tiempo de vida (en años) de estos televisores, tiene una función de densidad*

$$f_T(t) = \begin{cases} \frac{1}{63}(t-5)^2 & \text{si } 1 \leq t \leq c \\ 0 & \text{en otro caso} \end{cases}$$

Hallar la duración máxima de estos televisores, la duración esperada de los mismos.

Ejercicio 4 *En una muestra aleatoria de 142 mujeres de edad avanzada para un estudio sobre nutrición se registraron las siguientes variables: P : presión sanguínea C : concentración de colesterol en la sangre El coeficiente de correlación simple muestral resultó: $r_{EP} = 0,3332 r_{EC} = 0,5029 r_{PC} = 0,2495$ Se desea saber si existe alguna correlación lineal entre la presión sanguínea y la concentración de colesterol. Qué conclusión puede sacar ?*

Teórico 1 *Demostrar que la proporción muestral es un estimador insesgado de la probabilidad p de una variable binomial.*

Teórico 2 *Explicar el método de cuadrados mínimmos, definir e interpretar el coeficiente de Determinación e indicar cómo puede testearse la significación de un modelo de regresión lineal (indicando estadístico y distribución)*

PyE UTN Final HOJA N° FECHA 15-2-17

① Una reconocida marca de pilas lanza al mercado una nueva pila afirmando que la duración de estas es mayor a 120 horas. Una muestra de 23 pilas arrojó un promedio de duración de 122 horas y un desvío estándar de 4 horas. Suponiendo que la duración tiene distr. Normal.

a) Plantear las hipótesis del test y aplicarlo con nivel $\alpha = 0.05$

$$H_0: \mu = 120 \quad vs \quad H_1: \mu > 120$$

$n = 23$
$\bar{x} = 122$
$s = 4$

$$e_m = \frac{\bar{X} - 120}{4/\sqrt{23}} \sim N(0,1) \text{ bajo } H_0$$

Rechazo H_0 si $t_{obs} > t_{22,0.05}$

$$T_{obs} = \frac{122 - 120}{4} \cdot \sqrt{23} = 2,3979$$

$$t_{22,0.05} = 1,717$$

$> t_{obs} > t_{22,0,95}$ ∴ $\boxed{\text{Rechazo } H_0}$

b) Si el desvío estándar poblacional es conocido e igual a 4,5 con nivel de significación $\alpha = 0.05$, hallar el tamaño muestral tal que la prob. de equivocarse al tomar la decisión cuando la verdadera media es de 120,5 horas, sea menor a 0,02

$\sigma = 4,5 \quad \alpha = 0,05 \quad n = ?$ $\quad \mu = 120,5 \to H_0F \to$ equivocarse es aceptar H_0

$$P(\text{aceptar } H_0 \mid H_0F) \leq 0.02$$

$\to P(\text{error tipo II})$

Ahora se conoce $\sigma \to e_m = \frac{\bar{X} - 120}{4,5/\sqrt{n}} \sim N(0,1)$ bajo H_0

Rechazo H_0 si $z_{obs} > z_{0.95}$

$$P(\overbrace{\text{aceptar } H_0}^{\text{No rechazar } H_0} \mid H_0F) = P(z_{obs} \leq z_{0.95} \mid \mu = 120,5) =$$

$$= P\left(\frac{\bar{X} - 120}{4,5/\sqrt{n}} \leq 1,645 \,\middle|\, \mu = 120,5\right) = P\left(\frac{\bar{X} - 120,5 + 0,5}{4,5/\sqrt{n}} \leq 1,645\right) =$$

$$= P\left(\overbrace{\frac{\bar{X} - 120,5}{4,5/\sqrt{n}}}^{\sim N(0,1)} \leq 1,645 - \frac{0,5}{4,5} \cdot \sqrt{n}\right) = 0.02 = P(Z \leq -2,055)$$

$$\to -2,055 = 1,645 - \frac{0,5}{4,5}\sqrt{n} = \frac{7,4025 - 0,5\sqrt{n}}{4,5}$$

$$-9,2475 = 7,4025 - 0,5\sqrt{n} \to \sqrt{n} = \frac{7,4025 + 9,2475}{0,5} = 33,3$$

$$\to \sqrt{n} = 33,3 \to n = 1108,89 \to \boxed{n = 1109}$$

NOTA

③ Los televisores de plasma tiene una cierta vida útil, debido a la pérdida del gas que compone la pantalla. Suponiendo que el tiempo de vida (en años) de estos televisores, tienen una función de densidad.

$$f_T(t) = \begin{cases} \frac{(t-5)^2}{63} & \text{si } 1 \leq t \leq c \\ 0 & \text{en otro caso} \end{cases}$$

Hallar la duración máxima de estos televisores, la duración esperada de los mismos

$$F(c) = 1 = \int_1^c \frac{(t-5)^2}{63}\,dt = \frac{1}{63} \cdot \frac{(t-5)^3}{3}\Big|_1^c = \frac{(c-5)^3+64}{189} = 1 \rightarrow (c-5)^3 = 125$$

$$c-5 = 5 \rightarrow \boxed{c=10}$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_1^{10} \frac{x\,(x-5)^2}{63}\,dx = \frac{1}{63}\int_1^{10} x\,(x^2-10x+25)\,dx =$$

$$= \frac{1}{63}\int_1^{10} x^3-10x^2+25x\,dx = \frac{1}{63} \cdot \frac{1629}{4} = \frac{181}{28} = 6{,}46$$

$\boxed{\text{Duración máxima} = 10 \text{ años}}$ $\boxed{\text{Duración esperada} = 6{,}46 \text{ años}}$

④ En una muestra aleatoria de 142 mujeres de edad avanzada para un estudio sobre nutrición se registraron los sig. variables: P: presión sanguínea, C: concentración de colesterol en sangre. El coeficiente de correlación simple muestral resultó: $r_{EP} = 0{,}3332$, $r_{EC} = 0{,}5029$, $r_{PC} = 0{,}2495$. Se desea saber si existe alguna correlación lineal entre la presión sanguínea y la concentración de colesterol. ¿Qué conclusión puede sacar?

coef. correlación lineal en la muestra $= r_{PC} = 0{,}2495 = \rho$

Si $r = 0 \rightarrow$ no hay correlación. Cuanto más cercano a 1 o -1 mejor correlación. En este caso es 0,2495 $\rightarrow$ no hay buena correlación

NOTA

Final PyE UTN 15-2-17

② Sea X una v.a. continua que verifica $E(X)=8$, $V(X)=21$, $E(X^4)=7289$. Sea $Y=X^2$ una v.a. que representa "peso de un puma de cierta región de América del Sur"

1) Calcular la esperanza y varianza de Y

$E(Y) = E(X^2) \overset{c.A}{=} 85$ $\boxed{E(Y)=85}$

c.A $V(X) = E(X^2) - E(X)^2 = E(X^2) - 8^2 = 21 \rightarrow E(X^2) = 85$

$Y=X^2 \rightarrow Y^2 = X^4$

$V(Y) = E(Y^2) - \underbrace{E(Y)^2}_{85^2} = \underbrace{E(X^4)}_{x\text{-enunc.}=7289} - 85^2 = 7289 - 85^2 = 64$

$\boxed{V(Y) = 64}$

2) Si se eligen al azar 80 pumas para ser transportados ¿cuál es la prob. de sobrecarga sabiendo que la bodega tolera 6900 kg?

W: "peso de 80 pumas que pesan Y kg" $W = \sum_{i=1}^{80} Y$

$E(W) = E(\sum_{i=1}^{80} Y) = 80\,E(Y) = 6800 = \mu$

$V(W) = V(\sum_{i=1}^{80} Y) = 80\,V(Y) = 5120 = \sigma^2 \rightarrow \sigma = 71{,}55$

$\rightarrow W \overset{(a)}{\sim} N(6800, 71.55)$ x TCL

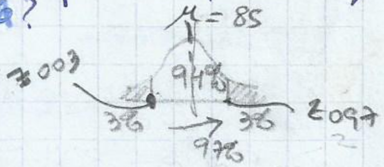
$Z = \dfrac{W-\mu}{\sigma} = \dfrac{W-6800}{71.55} \rightarrow Z \sim N(0,1)$

$P(W>6900) = P\left(\dfrac{W-6800}{71.55} > \dfrac{6900-6800}{71.55}\right) = P(Z>1{,}3976) = 1 - P(Z \le 1{,}3976) =$

calc $= 1 - 0{,}9189 = 0{,}0811$ $\rightarrow \boxed{P(\text{sobrecarga}) = 0{,}0811}$

3) ¿cuántos pumas hay que elegir en forma independiente para tener, al menos, un 94% de prob. de que el peso promedio esté entre 84 y 86?

$n = ?$ $\bar{Y} = \dfrac{\sum_{i=1}^{n} Y}{n}$ $\bar{Y} \overset{(a)}{\sim} N(85, 8/\sqrt{n})$



$P(84 < \bar{Y} < 86) = 0.94 = P(\bar{Y} < 86) - P(\bar{Y} \le 84) =$

x continuidad $= P(\bar{Y} \le 86) - P(\bar{Y} \le 84) = P\left(\dfrac{\bar{Y}-85}{8/\sqrt{n}} \le \dfrac{86-85}{8/\sqrt{n}}\right) - P\left(\dfrac{\bar{Y}-85}{8/\sqrt{n}} \le \dfrac{84-85}{8/\sqrt{n}}\right) =$

$\cong \overbrace{P\left(Z \le \underbrace{\dfrac{\sqrt{n}}{8}}_{z_{097}}\right)}^{0.97} - \underbrace{P\left(Z \le \overbrace{-\dfrac{\sqrt{n}}{8}}^{z_{003}}\right)}_{0.03} = 0.94$

$\rightarrow z_{003} = -1{,}88 \rightarrow -\dfrac{\sqrt{n}}{8} = -1{,}88 \rightarrow \sqrt{n} = 15{,}04 \rightarrow n = 226{,}20 \rightarrow \boxed{n = 227}$

teórico 1

Demostrar que la proporción muestral es un estimador insesgado de la prob. p de una variable binomial

$X \sim Be(p)$
$R(X) = \{0, 1\}$

Un buen estimador de p es $\bar{X}$ $\rightarrow \hat{p} = \bar{X}$

$Y \sim Bi(n, p)$ $E(Y) = np$ $V(Y) = np(1-p)$

$$\text{sesgo} = E(\hat{p}) - p = E(\bar{X}) - p = E\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) - p =$$
$$= \frac{\sum_{i=1}^{n} E(X_i)}{n} - p = \frac{n}{n} \overset{p}{E(X_i)} - p = p - p = 0$$

No sé si es esto lo que pide el enunciado

---

teórico 2

Explicar el método de cuadrados mínimos, definir e interpretar el coeficiente de Determinación e indicar cómo puede testearse la significación de un modelo de regresión lineal (indicando estadístico y distribución)

Cuadrados mínimos: es una técnica en la que, dados un conjunto de pares ordenados se busca una función continua para aproximar los datos con un criterio de mínimo error cuadrático

Con el modelo de regresión lineal se busca minimizar $\sum_{i=1}^{n} \left[Y_i - (\beta_0 + \beta_1 x_i)\right]^2$

Coeficiente de determinación: Es el porcentaje de variabilidad de Y que es explicado por el modelo de regresión lineal en X

$$r^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

Significación: $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$.

Estadístico:
$$T = \frac{b_1 - \overset{\beta_1}{0}}{S / \sqrt{S_{xx}}} \sim t_{n-2}$$

Rechazo si $|T_{obs}| > t_{n-2, \frac{\alpha}{2}}$

NOTA