20-12-17

Final Probabilidad y Estadística UTN 20/12/2017

Ej1)

Un estudio epidemiológico llevado a cabo sobre 800 adultos aparentemente sanos revelo que 210 presentaban evidencia serológica de infección toxoplasmosis.

a) Indicar un estimador puntual para la proporción poblacional de personas infectadas en esta población. Se trata de un estimador insesgado?
b) Estimar mediante un intervalo esta proporción con un nivel del 98%.
c) ¿Cuántos individuos deberían analizarse si se quiere disminuir el error de estimación del intervalo anterior en un 20%

Ej2) La función de densidad de la proporción de artículos reclamados de los producidos por cierta maquina es:

$$f(x) = \begin{cases} aX^{a-1} \; si \; x \in (0;1) \\ 0 \; en \; otro \; caso \end{cases}$$

a) Hallar el valor de a, sabiendo que el valor mediano de esta proporción es de $\frac{1}{\sqrt[3]{2}}$
b) Hallar el valor esperado, la varianza de esta proporción

Ej3) Se desea comparar las demandas medias de reparaciones de dos sucursales. Para ello se toman en forma aleatoria las demandas correspondientes a 12 días de cada una de ellas.

180	170	160	200	210	208	220	165	203	240	188	193
215	200	258	219	175	230	218	220	210	170	180	190

a) Considerando que las demandas tienen distribución normal y dispersiones iguales en ambas sucursales y utilizando un nivel de significación del 5%, testear las hipótesis de interés.
b) Definir y calcular el valor p

Ej4) Los errores que cometen dos mecanógrafas tienen distribución poisson con medias de 2 y 3 errores respectivamente. De un conjunto de hojas tipeadas por estas chicas se sabe que la 1era tipeo el doble de hojas que la 2da.

a) Cuál es la probabilidad de que no tenga errores? (1 hoja tipeada por cualquiera de las chicas)
b) Si no tiene errores. ¿Cuál es la probabilidad de que haya sido tipeda por la 1era mecanógrafa?

T1) Explique el método de cuadrados mínimos en el contexto de regresión lineal. Defina coeficiente de determinación e interprete su resultado

T2) Enuncie el Teorema central del límite y muestre un ejemplo de aplicación.

PyE UTN FINAL

HOJA N: FECHA 20-12-17

① Un estudio epidemiológico llevado a cabo sobre 800 adultos aparentemente sanos reveló que 210 presentaban evidencia serológica de infección toxoplasmosis.

a) Indicar un estimador puntual para la proporción poblacional de personas infectadas en esta población. ¿Se trata de un estimador insesgado?

X_i: "el iésimo adulto reveló evidencia serológica de toxoplasmosis"

$i \in [1, 800]$, $X_i \sim Be(p)$; $R(X) = \{0,1\}$, $\bar{X} = \frac{\sum_{i=1}^{800} X_i}{800} = \frac{210}{800}$

$\hat{p} = \bar{X}$, $\boxed{\hat{p} = 0,2625}$, $E(X) = p$

$$\text{Sesgo} = E(\hat{p}) - p = E(\bar{X}) - p = E\left(\frac{\sum_{i=1}^{800} X_i}{800}\right) - p = \frac{1}{800}\sum_{i=1}^{800} \overbrace{E(X_i)}^{p} - p =$$

$$= \frac{800p}{800} - p = p - p = 0 \rightarrow \boxed{\hat{p} \text{ es insesgado}}$$

b) Estimar mediante un intervalo para esta proporción con un nivel del 98%

$n = 800$, $\hat{p} = 0.2625$, $\alpha = 0,02 \rightarrow 1-\frac{\alpha}{2} = 0.99$ $\quad z_{0.99} = 2,3265$

$$IC_{0.98}(p) = \left[\hat{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] =$$

$$= \left[0.2625 - 2,3265\sqrt{\frac{0.2625(1-0,2625)}{800}}, 0.2625 + 2,3265\sqrt{\frac{0,2625(1-0,2625)}{800}}\right] =$$

$$\boxed{IC_{0.98}(p) = [0,2263 ; 0,2987]}$$

c) ¿Cuántos individuos deberían analizarse si se quiere disminuir el error de estimación del intervalo anterior en un 20%?

$$\text{error anterior} = 2,3265\sqrt{\frac{0.2625(1-0.2625)}{800}} = 0,03619 \rightarrow \text{error actual} \overset{80\% \text{ del anterior}}{=} 0,02895$$

$$\text{error actual} = 0.8 \text{ error anterior} = 0.02985 = 2,3265\sqrt{\frac{0.2625(1-0.2625)}{n}}$$

$$\frac{0,02985}{2,3265} = \frac{\sqrt{0.1936}}{\sqrt{n}} \rightarrow \sqrt{n} = 34,29 \rightarrow \boxed{n = 1176}$$

Sylvina

NOTA

② La función de densidad de la proporción de artículos reclamados de los producidos por cierta máquina es

$$f(x) = \begin{cases} a x^{a-1} & \text{si } x \in [0,1] \\ 0 & \text{en otro caso} \end{cases}$$

a) Hallar el valor de 'a' sabiendo que el valor mediano de esta proporción es $\frac{1}{\sqrt[3]{2}}$

Mediana → x tal que $f(x) = 0{,}5 \rightarrow f\left(\frac{1}{\sqrt[3]{2}}\right) = \int_0^{\frac{1}{\sqrt[3]{2}}} a x^{a-1}\, dx =$

$\rightarrow \left(\frac{1}{\sqrt[3]{2}}\right)^a = 0{,}5 \rightarrow \ln\left(\left(\frac{1}{\sqrt[3]{2}}\right)^a\right) = \ln(0{,}5) = a \ln\left(\frac{1}{\sqrt[3]{2}}\right) \rightarrow a = \frac{\ln(0{,}5)}{\ln\left(\frac{1}{\sqrt[3]{2}}\right)} = 3$

$$\boxed{a = 3}$$

b) Hallar el valor esperado y la varianza de esta proporción

- $E(X) = \int_0^1 x f(x)\, dx = \int_0^1 x \cdot 3x^2\, dx = \int_0^1 3x^3\, dx = \frac{3}{4} \rightarrow \boxed{E(X) = 0{,}75}$

- $V(X) = E(X^2) - E(X)^2 \underset{\text{C.A}}{=} \frac{3}{5} - \frac{9}{16} = \frac{3}{80} \rightarrow \boxed{V(X) = 0{,}0375}$

C.A: $E(X)^2 = 0{,}75^2 = \frac{9}{16}$

$E(X^2) = \int_0^1 x^2 f(x)\, dx = \int_0^1 3x^4\, dx = \frac{3}{5}$

HOJA N°

FECHA 20-12-17

PyE UTN Final

(3) Se desea comparar las demandas medias de reparaciones de dos sucursales. Para ello se toman en forma aleatoria las demandas correspondientes a 12 días de cada una de ellas

A	180	170	160	200	210	208	220	165	203	240	188	193
B	215	200	258	219	175	230	218	220	210	170	180	190

a) Considerando que las demandas tienen distribución normal y dispersiones iguales en ambas sucursales y utilizando un nivel de significación del 5% testear las hipótesis de interés.

$\sigma_A = \sigma_B$

$n_A = n_B = 12$

$\bar{X}_A = 194.75$

$\bar{X}_B = 207{,}08$

$\alpha = 0.05$

$S_A = 23{,}66$

$S_B = 25{,}40$

$$H_0: \underbrace{(\mu_B - \mu_A) = 0}_{\to \mu_A = \mu_B} \quad \text{vs} \quad H_1: \underbrace{(\mu_B - \mu_A) \neq 0}_{\to \mu_A \neq \mu_B}$$

$$e_m = \frac{(\bar{X}_B - \bar{X}_A) - (\mu_B - \mu_A)}{S_F \sqrt{\frac{1}{n_B} + \frac{1}{n_A}}}$$

$$S_F^2 = \frac{(n_B - 1) S_B^2 + (n_A - 1) S_A^2}{n_A + n_B - 2} \to S_F^2 = \frac{11 . (23{,}66 + 25{,}40)^2}{22} = 1203{,}44 \to S_F = 34{,}69$$

$$\to e_m = \frac{\bar{X}_B - \bar{X}_A}{34{,}69 \sqrt{\frac{1}{12} + \frac{1}{12}}} = \frac{\bar{X}_B - \bar{X}_A}{14{,}16} \sim t_{22}$$

Rechazo H_0 si $|T_{obs}| > t_{22,\, 0.025}$

$$T_{obs} = \frac{207{,}08 - 194{,}75}{14{,}16} = 0{,}8707$$

$$t_{22,\,0{,}025} = 2{,}074$$

$|T_{obs}| < t_{22,\,0.025}$ ∴ $\boxed{\text{No Rechazo } H_0}$

→ $\boxed{\text{No hay evidencias suficientes como para decir que son distintas}}$

b) Definir y calcular el valor p

p-valor es el mínimo nivel para el cual se rechaza la hipótesis nula

$$T_{obs} = 0{,}8707 \xrightarrow{\nu = 22} \boxed{0.15 \leq p\text{-valor} \leq 0.20}$$

$$\underbrace{t_{22,\,0.15}}_{0.858} \qquad \underbrace{t_{22,\,0.20}}_{1{,}061}$$

Sylvina

NOTA

④ Los errores que cometen dos mecanógrafas tienen distribución poisson con medias de 2 y 3 errores respectivamente. De un conjunto de hojas tipeadas por estas personas se sabe que la primera tipeó el doble de hojas que la segunda.

a) ¿Cuál es la prob. de que no tenga errores? (1 hoja tipeada por cualquiera de las mecanógrafas)

X: "cant de errores cometidos por la 1° mecanógrafa" $X \sim Po(2)$

Y: " " " " " " 2° mec." $Y \sim Po(3)$

A: "la mecanógrafa elegida al azar es la 1°" $P(A) + P(B) = 1$

B: " " " " " " " 2°" $P(B) = 2P(A)$

$$P(A) + P(B) = P(A) + 2P(A) = 3P(A) = 1 \rightarrow P(A) = 1/3 \qquad P(B) = 2/3$$

C: "cant de errores detectados en una hoja"

$$P(C=0) = P(C=0|A)P(A) + P(C=0|B)P(B) = P(X=0)\overbrace{P(A)}^{1/3} + P(Y=0)\overbrace{P(B)}^{2/3} =$$

$$= e^{-2}\frac{2^0}{0!}\cdot\frac{1}{3} + e^{-3}\frac{3^0}{0!}\cdot\frac{2}{3} = \frac{e^{-2}}{3} + \frac{2e^{-3}}{3} = 0,0783$$

$$\boxed{P(C=0) = 0,0783}$$

b) Si no tiene errores ¿cuál es la prob. de que haya sido tipeada por la primera mecanógrafa?

$$P(A|C=0) = \frac{P(C=0|A)P(A)}{P(C=0|A)P(A) + P(C=0|B)P(B)} = \frac{P(X=0)P(A)}{0,0783} = \frac{\frac{e^{-2}}{3}}{0,0783}$$

$$\boxed{P(A|C=0) = 0,5761}$$

NOTA

PyE UTN

HOJA N°

FECHA 20-12-17

T1 Explique el método de cuadrados mínimos en el contexto de regresión lineal. Defina coeficiente de determinación e interprete su resultado

Consiste en explicar una de las variables en función de la otra a través de un determinado tipo de función de forma que la función de regresión se obtiene ajustando las observaciones a la función elegida, mediante el método de cuadrados mínimos

→ Busco minimizar $\sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 X_i)]^2$

donde X_i e Y_i son v.a. y el modelo de regresión lineal es

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Coeficiente de determinación:

Es el porcentaje de la variabilidad de Y que es explicado por el modelo de regresión lineal en X

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

Si r es 0 significa que la variable predictoria no tiene NULA capacidad predictiva

Cuando mayor sea r mejor será la predicción $0 \leq r \leq 1$

T2 Enunciel el teorema central del límite y muestre un ejemplo de aplicación

Sean $X_1, X_2 \ldots$ una sucesión de v.a. i.i.d (indep idénticamente distribuidas) tales que $E(X_i) = \mu$ y $V(X_i) = \sigma^2$

Sea $X = X_1 + X_2 + \ldots + X_n$ con n suficientemente grande.

Entonces, bajo ciertas condiciones grales.:

$$\frac{X - n\mu}{\sigma\sqrt{n}} \longmapsto Z \text{ donde } Z \sim N(0,1)$$

ej.:

Un cajón contiene 49 manzanas. El valor esperado de la manzana es de 150gr. ¿Cuál es la prob de que el cajón pese menos de 7,25 kg? y varianza $= 225\,gr^2$

$n = 49$ $\quad X = \sum_{i=1}^{49} X_i$, $E(X_i) = 150$ $\quad V(X) = 225 \rightarrow \sigma = 15$

$$\frac{X - \overbrace{49 \times 150}^{7350}}{15 \times 7} = Z \overset{a}{\sim} N(0,1) \quad \text{(si no se conoce la distr. → aproxima a } N(0,1))$$

$$P(X < 7250) \overset{\text{continua}}{=} P(X \leq 7250) = P\left(\frac{X - 7350}{105} \leq \frac{7250 - 7350}{105}\right) \cong P(Z \leq -0{,}9523) =$$

$$= 0{,}1705 \rightarrow \boxed{P(X < 7250) = 0{,}1705}$$

Sylvane

NOTA