27-9-17

Universidad Tecnológica Nacional Facultad Regional Buenos Aires Examen Final de Probabilidad y Estadística - 27 de septiembre 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 *La función de distribución del contenido de un reservorio de agua es :*

$$F_X(x) = \begin{cases} 0 & si & x \in (-\infty, 0) \\ \frac{x^3}{\theta} & si & x \in [0;2) \\ 1 & si & x \in [2;+\infty) \end{cases}$$

a. *Hallar el valor de θ y el valor medio de la variable.*

b. *Hallar la probabilidad de que en dos mediciones independientes del contenido del reservorio ambas superen la mediana.*

Ejercicio 2 *Al lanzar 5000 veces una moneda al aire salieron 3000 caras. ¿Se puede aceptar, con un nivel de significación del 0.04, que la moneda no está cargada? Plantee las hipótesis a testear y proponga un test de nivel adecuado para este problema. Indique claramente la zona de rechazo y la decisión.*

Ejercicio 3 *Una investigación de la relación entre flujo de tránsito X (miles de automóviles en 24hs) y el contenido de plomo de la corteza de los árboles cerca de la autopista Y ($\mu g/g$ de peso en seco) produjo los datos siguientes:*

x	8.3	8.3	12.1	12.1	17	17	17	24.3	24.3	24.3	33.6
y	227	312	362	521	640	539	728	945	915	810	1340

a. *Hallar un modelo lineal para estimar el contenido de plomo de la corteza en función del flujo de automóviles.*

b. *Analizar la significación del modelo propuesto al 5%.*

Ejercicio 4 *Se analizó una muestra de 12 piezas de pan blanco de cierta marca y se determinó el porcentaje de carbohidratos contenido en cada una de las piezas, obteniéndose los siguientes valores:*

76.93 76.88 77.07 76.68 76.39 75.09 77.67 76.88 78.15 76.50 77.16 76.42

a) Estimar puntualmente el contenido medio de carbohidratos de las piezas de pan de esta marca, sabiendo que su distribución es normal.

b) Estimar mediante un intervalo de 95% de confianza el contenido medio de carbohidratos de estos panes.

Teórico 1 *Sean $S_1, \ldots, S_n$ n eventos independientes dicotómicos, tales que $P(S_i = 1) = p_i$. Expresar en función de p_i la probabilidad a) de que todos ocurran b) que ninguno ocurra y c) al menos uno ocurra. Si $p_i = p$ para todo i, qué distribución tienen X: número de eventos que ocurrieron.*

Teórico 2 a. *Definir estimador. Señalar dos propiedades deseables y ejemplificar con algún estimador que las tenga.*

b. *Explicar el concepto de p valor.*

P y E - final 27-9-17

① La función de distribución del contenido de un reservorio de agua es:

$$F_X(x) = \begin{cases} 0 & \text{si } x \in (-\infty; 0) \\ \frac{x^3}{\theta} & \text{si } x \in [0; 2) \\ 1 & \text{si } x \in [2; +\infty) \end{cases}$$

a) Hallar el valor de θ y el valor medio de la variable

$$F_X(2) = 1 = \frac{2^3}{\theta} = \frac{8}{\theta} \rightarrow \boxed{\theta = 8}$$

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx \quad , \quad F(x) = \int_{-\infty}^{x} f(t) dt \rightarrow f(t) = F'(t) = \boxed{\frac{3}{8} t^2 = f(t)}$$

$$E(X) = \int_0^2 x \frac{3}{8} x^2\, dx = \boxed{\frac{3}{2} = E(X)}$$

b) Hallar la prob. de que en dos mediciones independientes el contenido del reservorio ambas superen la mediana

q: La mediana es el valor de x en donde $F(q) = 0.50 \rightarrow \underbrace{P(X \leq q)}_{0.5} = 1 - \underbrace{P(X > q)}_{0.5}$

$\rightarrow p = P(X > q) = 0.5$

Y: "cantidad de mediciones que superan la mediana, de un total de 2"

$Y \sim Bi(2, 0.5)$

$$\rightarrow P(X=2) = \binom{2}{2} 0.5^2\, 0.5^0 = 1/4$$

$$\boxed{P(X=2) = 0.25}$$

② Al lanzar 5000 veces una moneda al aire salieron 3000 caras, ¿Se puede aceptar, con un nivel de significación del 0.04 que la moneda no está cargada? Plantee las hipótesis a testear y proponga un test de nivel adecuado para este problema. Indique claramente la zona de rechazo y la decisión.

$n = 5000$

$\hat{p} = \frac{3000}{5000} = 0.6$

$\alpha = 0.04$

$H_0 : p = 0.50 \quad vs \quad H_1 : p \neq 0.50$

$$e_m = \frac{p - 0.5}{\sqrt{\frac{0.5\,(0.5)}{5000}}} \sim N(0,1) \text{ bajo } H_0$$

$$e_m = \frac{p - 0.5}{\sqrt{2}/200}$$

$$\text{Rechazo } H_0 \text{ si } |Z_{obs}| > \overbrace{Z_{1-\frac{0.04}{2}}}^{Z_{098}}$$

$$Z_{obs} = \frac{0.6 - 0.5}{\sqrt{2}/200} = 14,1421$$

$Z_{098} = 2,055$

$Z_{obs} > Z_{098} \rightarrow$ Rechazo H_0

$\boxed{\text{No existen evidencias para asegurar que la moneda no está cargada}}$

③ Una investigación de la relación entre flujo de tránsito X (miles de automóviles en 24hs) y el contenido de plomo de la corteza de los árboles cerca de la autopista Y (Pg/g de peso en seco) produjo los sig. datos

X	8.3	8.3	12.1	12.1	17	17	17	24.3	24.3	24.3	33.6
Y	227	312	362	521	640	539	728	945	915	810	1340

a) Hallar un modelo lineal para estimar el contenido de plomo de la corteza en función del flujo de automóviles.

$a = b_0 = -61,17$

$b = b_1 = 40,40$

$$\boxed{\hat{Y} = -61,17 + 40,40\,X}$$

b) Analizar la significación del modelo propuesto al 5%

$n = 11$

$\alpha = 0.05$

$\frac{b_1 - 0}{S/\sqrt{S_{XX}}}$

• $H_0 : \beta_1 = 0 \quad vs \quad \beta_1 \neq 0$

$$S^2 = \frac{1}{n-2}\left(S_{YY} - \frac{S_{XY}^2}{S_{XX}}\right) = \frac{1}{9}\cdot\left(1066965,64 - \frac{25179,85^2}{623,22}\right) = 5514,65 \rightarrow S = 74,26$$

• $\frac{b_1}{74,26/24,95} \sim t_9 \rightarrow$ Rechazo H_0 si $|t_{obs}| > t_{9,\,0.025}$

$$\left\{\begin{array}{l} T_{obs} = \frac{40,40}{74,26/24,95} = 13,57 \\ t_{9,\,0.025} = 2,262 \end{array}\right\} T_{obs} > t_{9,\,0.025} \rightarrow \boxed{\text{Rechazo } H_0}$$

PyE final 27-9-17

(4) Se analizó una muestra de 12 piezas de pan blanco de cierta marca y se determinó el porcentaje de carbohidratos contenido en cada una de las piezas, obteniéndose los sig. valores.

76.93 76.88 77.07 76.68 76.39 75.09 77.67 76.88 78.15
76,50 77.16 76,42

a) estimar puntualmente el contenido medio de carbohidrato de las piezas de pan de esta marca, sabiendo que su distr. es normal

$\boxed{\bar{X} = 76{,}82}$ $\hat{\mu} = \bar{X}$

b) Estimar mediante un intervalo de 95% de confianza el contenido medio de carbohidratos de estos panes

$IC_{0.95}(\mu) = \left[\bar{X} - t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} , \bar{X} + t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right] =$

σ desc.

$S^2 = 0{,}7476$

$\hookrightarrow S = 0.865$

$t_{11, 0.025} = 2{,}201$

$= \left[76.82 - 2{,}201 \cdot \frac{0.865}{\sqrt{12}} ; 76.82 + 2{,}201 \times \frac{0.865}{\sqrt{12}}\right]$

$\Rightarrow \boxed{IC_{0.95}(\mu) = [76{,}27 ; 77{,}37]}$

Teorico 1 | Sean $S_1, \dots S_n$ n eventos independientes dicotómicos tales que $P(S_i = 1) = p$. Expresar en función de p la probabilidad de: bernoullies

a) de que todos ocurran: $P(X = n) = \binom{n}{n} p^n (1-p)^0 = \boxed{p^n = P(X=n)}$

b) que ninguno ocurra: $P(X=0) = \binom{n}{0} p^0 (1-p)^n = \boxed{(1-p)^n = P(X=0)}$

c) al menos uno ocurra: $P(X \geq 1) = 1 - P(X<0) = 1 - P(X=0) = \boxed{1 - (1-p)^n = P(X \geq 1)}$

Si $p_i = p$. $\forall i$ qué distribución tiene: X: 'número de eventos que ocurrieron'

$X \sim Bi(n, p)$

teórico 2

a) Definir estimador. Señalar dos propiedades deseables y ejemplificar con algún estimador que las tenga

Un estimador es una variable aleatoria que se utiliza para estimar (aproximar) el valor de un parámetro de otra variable.

prop. 1: que sea insesgado, $E(X) = \mu$, $V(X) = \sigma^2$

$\hat{\mu} = \bar{X}$ para estimar μ

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

$$= \frac{\sum_{i=1}^{n} E(X_i)}{n} - \mu = \frac{n}{n} \underbrace{E(X_i)}_{\mu} - \mu = \mu - \mu = 0$$

insesgado.

Prop 2: que sea consistente.

$$\hat{\mu} \xrightarrow[n \to \infty]{} \mu \quad ; \quad \bar{X} \xrightarrow[n \to \infty]{} \mu$$

b) Explicar el concepto de p-valor

- Es la probabilidad de obtener el resultado que se obtiene si se supone que Ho es cierta.

  Si el valor de p es inferior al nivel de significación lo más verosímil es que la hipótesis de partida sea falsa

- Es el nivel de significación más bajo en el que el valor observado de la muestra es significativo. Es el nivel de significación más pequeño que conduce al rechazo de Ho.