2016

Ejercicio 3

En años recientes los granjeros suecos fumigaron sus campos sembrados de cereales con un fungicida que contenía metil de mercurio. Para tener una estimación del grado de contaminación inducido a productos comestibles, se realizó un estudio sobre el nivel de mercurio de los huevos producidos en Suecia. Para tal fin se tomó una muestra aletoria de 1600 huevos y se observó que 940 de ellos estaban contaminados (esto es, tenían un nivel de metil de mercurio superior al máximo tolerado). En lo que sigue, denotamos por p a la proporción de huevos contaminados.

a) Construya un intervalo de confianza 90 % aproximado para p.

b) Realice una prueba de hipótesis al nivel $\alpha = 0,10$ para decidir entre las siguientes hipótesis:

$$\begin{cases} H_0: & p = 0,6 \\ H_1: & p < 0,6 \end{cases}$$

c) Sabiendo que el verdadero valor de p es 0,55, calcule la potencia de la prueba.

$$X_i \sim Ber(p) \Rightarrow \overline{X}_n \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

$$\sigma = \frac{p(1-p)}{n}$$

$$p = \frac{940}{1600}$$

Ejercicio 7

Se toma una muestra aleatoria de n habitantes de una ciudad muy grande, en la que una proporción p de personas padecen cierta enfermedad.

1. Si n = 400 y se encuentran 165 personas enfermas, estimar p y construir un intervalo de confianza 95 % para p.
2. Hacer una prueba de hipótesis para decidir si $p = 0{,}40$

$$\begin{cases} H_0: & p = 0{,}4 \\ H_1: & p \neq 0{,}4 \end{cases}$$

$$X_j = \begin{cases} 1 & \text{si está enferma} \quad (p) \\ 0 & \text{si no.} \quad (1-p) \end{cases}$$

$$X_j \sim Ber(p)$$

$$\Rightarrow \overline{X_n} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

$$\hat{p} = \frac{165}{400} = 0{,}4125 \rightarrow p \text{ observado}$$

La región crítica se determina para que:

$$P_{H_0}((x_1, \ldots, x_n) \in RC) = \alpha$$

$\rightarrow p = 0{,}4$

$$\{ |\overline{X_n}| \geq 0{,}4 + K \}$$

Ejercicio 8

Se considera la muestra i.i.d. $X_1, \dots, X_{200}$ proveniente de una distribución $U(0,b)$, tal que: $\sum_{i=1}^{200} X_i = 928{,}68$ y $\sum_{i=1}^{200} X_i^2 = 5726{,}77$.

$S_n \bar{X}_n$ $\alpha = 0{,}10$ $\frac{b}{2}$

a) Encuentre un intervalo de confianza asintótico al nivel 0.90 para $m = E(X)$.

b) Si los datos provienen de una distribución $U(0,b)$, trabajando con $\alpha = 0{,}05$ encuentre la región de rechazo del test $\begin{cases} H_0: & b = 10 \\ H_1: & b < 10 \end{cases}$
(Sugerencia: recuerde que si $X \sim U(a,b)$ entonces $E(X) = \frac{a+b}{2}$).

Chat

Si $X \sim U(a,b) \Rightarrow E(X) = \frac{a+b}{2}$

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2\right)$$

La varianza muestral

Si $X \sim U(0,b) \Rightarrow \overline{X}_n \sim N\left(\frac{b}{2}, \frac{S_n}{n}\right)$

$\sum_{i=1}^{200} X_i = 928{,}68$ y $\sum_{i=1}^{200} X_i^2 = 5726{,}77$.

$$S_n = \sqrt{\frac{1}{900}\left(\sum_{i=1}^{200} X_i^2 - 200\left(\frac{1}{n}\sum_{i=1}^{200} X_i\right)^2\right)}$$

$5726{,}77$ $\quad 928{,}68 \quad \frac{1}{200}$

$$= \sqrt{\frac{1}{900}\left(5726{,}77 - 4312{,}23\right)}$$

$$= \sqrt{\frac{1414{,}53}{900}} = \sqrt{1{,}571} = 1{,}253$$

$$\overline{X}_n \sim N\left(\frac{b}{2}, \frac{1{,}253}{200}\right)$$

σ

$$P\left(|\overline{X}_n| \leq \frac{b}{2} + k\right) = 0{,}9$$

$$P\left(-\frac{b}{2} - k \leq \overline{X}_n \leq \frac{b}{2} + k\right)$$

$\vdots$

Normalizar y encontrar k. $\overline{X}_n \to \frac{b}{2}$

b) Si los datos provienen de una distribución $U(0,b)$, trabajando con $\alpha = 0{,}05$ encuentre la región de rechazo del test $\begin{cases} H_0: & b = 10 \\ H_1: & b < 10 \end{cases}$

(Sugerencia: recuerde que si $X \sim U(a,b)$ entonces $E(X) = \frac{a+b}{2}$)

$$RC = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : 2\overline{X}_n < 10 - k\}$$

$b = 10$

$$P_{H_0}\left(\left((x_1, \ldots, x_n) \in R_c\right)\right) = 0{,}05$$

$$P(2\overline{X}_n < 10 - k)$$

$$P\left(\overline{X}_n < \frac{10-k}{2}\right)$$

$X \sim U(a,b)$

$E(X) = \frac{a+b}{2}$

$\sigma^2 = \frac{(b-a)^2}{12}$

$$P\left(\underbrace{\frac{\left(\overline{X}_n - \frac{10}{2}\right) 12\sqrt{200}}{100}}_{\sim N(0,1)} < \left(\frac{10-k}{2} - \frac{10}{2}\right)\frac{12\sqrt{200}}{100}\right)$$

$$P\left(Z < -12k \cdot \frac{\sqrt{200}}{200}\right) \qquad \frac{1}{\sqrt{200}}$$

$$P\left(Z < \frac{-12k}{\sqrt{200}}\right) = 0{,}05$$

Kolmogorov - Smirnov

$X_1, \dots, X_n$ MAS , $X \sim F_x$ desconocida.

$$\begin{cases} H_0: & F_x = F_0 \\ H_1: & \text{No } H_0 \end{cases}$$

Conocemos F_0

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\}$$

$$\begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{si no} \end{cases}$$

$$RC = \left\{ \sup_{x\in\mathbb{R}} \left| F_n^*(x) - F_0(x) \right| \geq c \right\}$$

Constante a determinar.

$$P\big((X_1, \dots, X_n) \in RC\big) = \alpha$$

Obs:

$$\sup_{x\in\mathbb{R}} \left| F_n^*(x) - F_0(x) \right| = \max\left\{ \max_{1\leq i\leq n} \left| \frac{i}{n} - F_0(X_i^*) \right|, \max_{1\leq i\leq n} \left| \frac{i-1}{n} - F_0(X_i^*) \right| \right\}$$

Ejercicio 1

Para la siguiente muestra

0.073	0.021	0.162	0.094	0.303
0.018	0.08	0.061	0.19	0.079

10

Realizar la prueba de Kolmogorov y Smirnov para ver si es razonable afirmar que los datos tienen distribución exponencial de parámetro $\lambda = 2$.

Ejercicio 2

$$\begin{cases} H_0: & F = \begin{cases} 2e^{-2x} & \text{si } x \geq 0 \\ 0 & \text{si } x \leq 0 \end{cases} \\ H_1: & \text{No } H_0 \end{cases}$$

$$\sup_{x\in\mathbb{R}} \left|F_n^*(x) - F_0(x)\right| = \max\left\{ \max_{1\leq i\leq n} \left|\frac{i}{n} - F_0(X_i^*)\right|, \max_{1\leq i\leq n} \left|\frac{i-1}{n} - F_0(X_i^*)\right| \right\}$$

0.073	0.021	0.162	0.094	0.303
0.018	0.08	0.061	0.19	0.079

i	X_i^*	$\left\lvert \frac{i}{n} - 2e^{-2X_i^*} \right\rvert$	$\left\lvert \frac{i-1}{n} - 2e^{-2X_i^*} \right\rvert$
1	0,018	$\left\lvert \frac{1}{10} - 2e^{-0,036} \right\rvert$	
2	0,021	$\left\lvert \frac{2}{10} - 2e^{-0,042} \right\rvert$	
3	0,061	$\left\lvert \frac{3}{10} - 2e^{-0,122} \right\rvert$	
4	0,073	$\left\lvert \frac{4}{10} - 2e^{-0,146} \right\rvert$	
5	0,079	$\left\lvert 5/10 - 2e^{-0,158} \right\rvert$	
6	0,08		
7	0,094		
8	0,162		
9	0,19		
10	0,303		

i	x_i^*	$\left\lvert \frac{i}{n} - 2e^{-2x_i^*} \right\rvert$	$\left\lvert \frac{i-1}{n} - 2e^{-2x_i^*} \right\rvert$
1	0,018	1,82928059	1,92928059
2	0,021	1,71773956	1,81773956
3	0,061	1,47029674	1,57029674
4	0,073	1,32831541	1,42831541
5	0,079	1,20769956	1,30769956
6	0,08	1,10428758	1,20428758
7	0,094	0,95722941	1,05722941
8	0,162	0,64650048	0,74650048
9	0,19	0,46772282	0,56772282
→ 10	0,303	0,09105725	0,19105725

$$\sup_{x \in \mathbb{R}} \{ |F_n^*(x) - F_0(x)| \} = 1,929$$

$$RC = \{ (x_1, \dots, x_n) \in \mathbb{R}^n : 1,929 \geq c \}$$

Elegimos $\alpha = 0,05$

Tabla 8 Contraste de Kolmogorov-Smirnov

Valores críticos de $D = |F_n(x) - F(x)|$ donde $F_n(x)$ es la distribución muestral de tamaño n y $F(x)$ la distribución teórica.

Tamaño muestral n	Nivel de significación 0,20	0,15	0,10	0,05	0,01
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490

$$RC = \{ \sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| \geq 0,410 \}$$

En nuestro caso $\sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)| = 1,3\ldots > 0,410$

entonces rechazo H_0.