Clase 6

jueves, 6 de abril de 2023 18:48

Intervalos de confianza

Ejercicio 6

Se arroja 50 veces una moneda con probabilidad p de salir cara. Hallar un intervalo de confianza asintótico de nivel $\underbrace{0.95}_{1-\alpha=0,95}$ para p basado en la observación x=50. $=\sum_{i=1}^{50} x_i$

$$X_i = \begin{cases} 1 \to \text{si sale cara} \\ 0 \to \text{si sale ceca} \end{cases}, \quad X_i \sim Be(p), \quad i=1;2;\dots;50$$

Estimador de p: $\hat{p} = \overline{X} = \dfrac{\sum_{i=1}^{50} X_i}{50}$, $X = \sum_{i=1}^{50} X_i \sim Bi(50;p)$

Pivote:

$$U = \frac{\overline{X} - p}{\sqrt{p(1-p)}} \cdot \sqrt{50} \overset{(a)}{\sim} N(0;1) \quad \Big| \quad Z = \frac{X - \mu}{\sigma}$$

Normal

$\mu \to E[x]$

Por Teorema Central del límite

$$E[\overline{X}] = E\left[\frac{\sum_{i=1}^{50} X_i}{50}\right] = \frac{1}{50}\sum_{i=1}^{50} E[X_i] = \frac{1}{\cancel{50}} \cancel{50} \cdot E[X_1] = p$$

$$V[\overline{X}] = \frac{p\cdot(1-p)}{50} \longrightarrow \sigma_{\overline{x}} = \sqrt{\frac{p(1-p)}{50}}$$

$Z \sim N(0;1)$

$$P(a < U < b) \simeq 1-\alpha \Rightarrow P\left(z_{0,025} < \frac{\overline{X}-p}{\sqrt{p(1-p)}} \cdot \sqrt{50} < z_{0,975}\right) \simeq 0,95$$

$$I(\underline{X}) = \left(\overline{X} - z_{0,975} \sqrt{\frac{p(1-p)}{50}} \, , \, \overline{X} + z_{0,975} \cdot \sqrt{\frac{p(1-p)}{50}}\right)$$

$$U = \frac{\overline{X} - p}{\sqrt{\underset{\overline{X}}{\hat{p}}(1-\underset{\overline{X}}{\hat{p}})}} \cdot \sqrt{50} \overset{(a)}{\sim} N(0,1)$$

$\to$ Teorema de Slutsky

- De la muestra se obtuvo $x = 50 \Rightarrow \hat{p} = \overline{x} = \frac{50}{50} = 1$

$$I.C.(p) = \left(1 - 1,96 \cdot \sqrt{\frac{1\cdot(1-1)}{50}} \, ; \, 1 + 1,96 \sqrt{\frac{1\cdot(1-1)}{50}}\right)$$

$I.C.(p) = (1;1)$ con 95% de confianza de nivel asintótico.

Test de hipótesis

Ejercicio 1

Supongamos que en las especificaciones de procedimientos de una planta de energía nuclear se establece que la resistencia media de soldadura debe superar 100lb/plg. Supongamos que somos el director del equipo de inspección del ente regulador estatal que debe determinar si la planta cumple con las especificaciones. e planea seleccionar una muestra al azar de soldaduras y realizar pruebas en cada una de ellas.

1. ¿Cuáles son las hipótesis a testear?
2. Explicar que significa en este contexto el error de tipo I y el de tipo II, y discutir cuáles son las consecuencias de cometer cada tipo de error

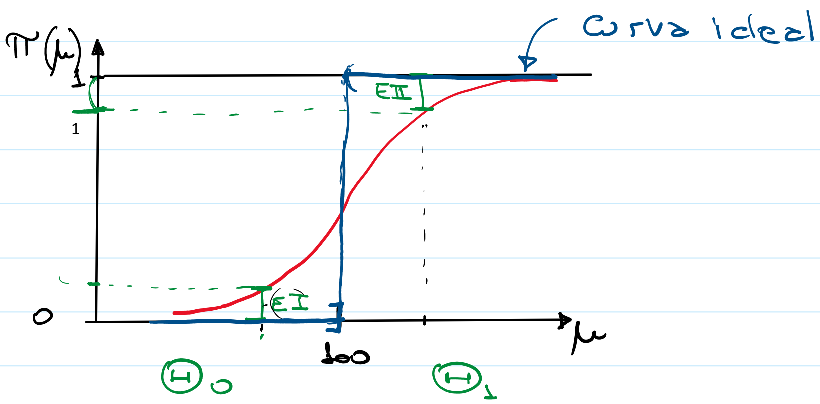
X_i: resistencia de la soldadura i, $i=1,2,\ldots,n$

$H_0: \mu \leq 100$ vs. $H_1: \mu > 100$

Error tipo I: decir que la planta cumple con las especificaciones de las soldaduras ($\mu > 100$) cuando en realidad no cumple.

Error tipo II: decir que la planta no cumple con las especificaciones cuando en realidad cumple.

Curva de potencia



Ejercicio 3

Se realiza un estudio para estimar la proporción de residentes en una ciudad que están a favor de la construcción de una planta de energía nuclear. En una muestra de 100 personas se observó que la proporción de individuos que se encuentran a favor fue de 0.62

1. ¿Puede decirse con un nivel de significación de 0.01 que la mayor parte de la población está a favor de la construcción de la planta nuclear?
2. Hallar el p-valor.

$$X_i = \begin{cases} 1 & \text{si la persona } i \text{ dice estar a favor} \\ 0 & \text{si la persona } i \text{ dice estar en contra} \end{cases}, \quad X_i \sim Be(p), \quad i = 1, 2, \dots, 100$$

$$H_0: p \leq 0{,}5 \quad \text{vs.} \quad H_1: p > 0{,}5$$

Estimador de p: $\hat{p} = \overline{X} = \dfrac{\sum_{i=1}^{100} X_i}{100}$

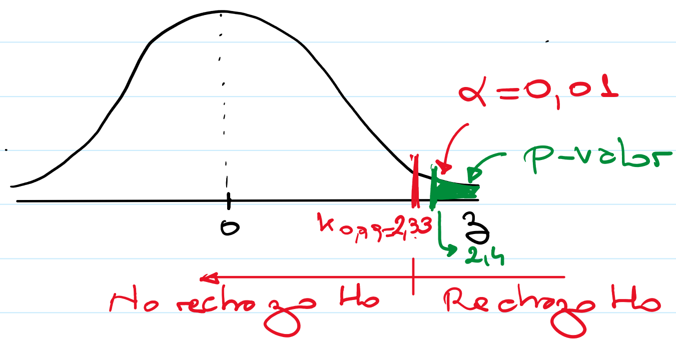
Estadístico de prueba: $\dfrac{\overline{X} - p}{\sqrt{p(1-p)}} \cdot \sqrt{100} \overset{(a)}{\sim} N(0,1)$

Por T.C.L.

Regla de decisión:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \dfrac{\overline{x} - 0{,}5}{\sqrt{0{,}5 \cdot 0{,}5}} \cdot \sqrt{100} > k_\alpha = k_{0{,}01} = 2{,}33 \\ 0 & \text{en caso contrario} \end{cases}$$

Bajo Ho verdadera.



- De la muestra se obtuvo: $\overline{x} = 0{,}62$, entonces:

$$\frac{0{,}62 - 0{,}5}{\sqrt{0{,}5 \cdot (1 - 0{,}5)}} \cdot \sqrt{100} = 2{,}4 > k_{0{,}01} = 2{,}33 \Rightarrow \text{se rechaza Ho.}$$

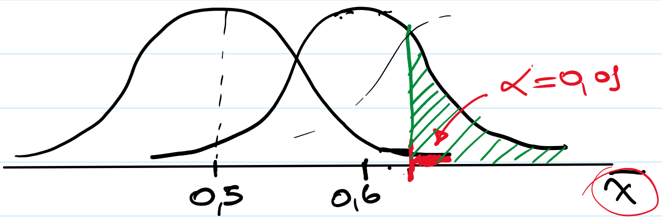
$$\text{p-valor} = P_{p=0{,}5}(\overline{X} > 0{,}62) = P_{p=0{,}5}\left(\frac{\overline{X} - 0{,}5}{\sqrt{0{,}5 \cdot (1-0{,}5)}} \cdot \sqrt{100} > \frac{0{,}62 - 0{,}5}{\sqrt{0{,}5 \cdot (1 - 0{,}5)}} \sqrt{100}\right) =$$

$$= P_{p=0{,}5}\left(\frac{\overline{X} - 0{,}5}{\ \ } \cdot \sqrt{100} > 2{,}4\right) = 0{,}0082$$

$$= P_{p=0,5}\left(\frac{\overline{X}-0,5}{\sqrt{0,5\cdot 0,5}}\cdot\sqrt{100} > 2,4\right) = 0,0082$$

$$\frac{\overline{X}-0,5}{\sqrt{0,5\cdot 0,5}}\cdot\sqrt{100} \to Z$$

- Calcular la potencia del test si $p = 0,6$



$$\pi(0,6) = P_{p=0,6}\left(\delta(\underline{X})=1\right) = P_{p=0,6}\left(\frac{\overline{X}-0,5}{\sqrt{0,5\cdot 0,5}}\cdot\sqrt{100} > 2,33\right) =$$

$$= P_{p=0,6}\left(\overline{X} > \frac{2,33\cdot\sqrt{0,5\cdot 0,5}}{10} + 0,5\right) =$$

$$= P_{p=0,6}\left(\frac{\overline{X}-0,6}{\sqrt{0,6\cdot 0,4}}\cdot\sqrt{100} > \left(\frac{2,33\cdot\sqrt{0,5\cdot 0,5}}{10} + 0,5 - 0,6\right)\cdot\frac{\sqrt{100}}{\sqrt{0,6\cdot 0,4}}\right) =$$

$$= P_{p=0,6}\left(\frac{\overline{X}-0,6}{\sqrt{0,6\cdot 0,4}}\sqrt{100} > 0,33\right) = 1 - \underbrace{\phi(0,33)}_{P(Z<0,33)} =$$

$$= 1 - 0,6293 = 0,3707$$

$$\frac{\overline{X}-0,6}{\sqrt{0,6\cdot 0,4}}\sqrt{100} \to Z$$

Ejercicio 2

Se tiene una m.a. de tamaño n de una población uniforme en el intervalo $(0,\theta)$.

1. Diseñar un test de hipótesis para decidir si θ es mayor a 2.5 con un nivel de significación de 0.05.
2. Suponer θ=3, n=20, simular la m.a. y decidir en base a ella.
3. Hallar el p-valor.

$$X_i \sim \mathcal{U}(0;\theta), \quad i = 1;2;\dots;n$$

$$H_0: \theta \leq 2,5 \quad \text{vs.} \quad H_1: \theta > 2,5$$

Estimador de máxima verosimilitud de θ: $\hat{\theta} = U = \max\{X_i\}_{i=1}^{n}$

$$F_U(u) = P(U \leq u) = P(X_1 \leq u, X_2 \leq u, \dots; X_n \leq u) =$$

$$= P(X_1 \leqslant u) \cdot P(X_2 \leqslant u) \cdot \ldots \cdot P(X_n \leqslant u) =$$

$$= \left[F_{X_1}(u)\right]^n = \left(\frac{u}{\theta}\right)^n \mathbb{1}\{0 < u < \theta\}$$

$$F_U(u) = \left(\frac{u}{\theta}\right)^n \mathbb{1}\{0 < u < \theta\}$$

Propongo un estadístico de prueba: $T = \dfrac{\underset{U}{\underbrace{\text{máx}\{X_i\}_{i=1}^n}}}{\theta}$

$$F_T(t) = P(T \leqslant t) = P\left(\frac{\text{máx}\{X_i\}_{i=1}^n}{\theta} \leqslant t\right) =$$

$$= P(U \leqslant \theta t) = F_U(\theta t) = \left(\frac{\cancel{\theta} \cdot t}{\cancel{\theta}}\right)^n = t^n \mathbb{1}\{0 < t < 1\}$$

$$\underbrace{\delta(\underline{X})}_{\text{Bajo } H_0 \text{ verdadera}} = \begin{cases} 1 & \text{si } \dfrac{\text{máx}\{x_i\}_{i=1}^n}{2{,}5} > k_{0{,}05} = t_{0{,}05} = \sqrt[n]{0{,}95} \Rightarrow \text{se rechaza } H_0 \\ 1 & \text{si } \dfrac{\text{máx}\{x_i\}_{i=1}^n}{2{,}5} < k_{0{,}05} \end{cases}$$

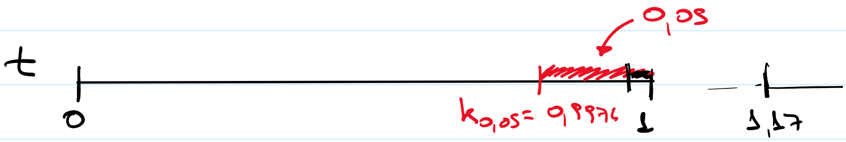
$$P(T > t_{0{,}05}) = 0{,}05 \Rightarrow 1 - F_T(t_{0{,}05}) = 1 - t_{0{,}05}^n = 0{,}05 \Rightarrow$$

$$\Rightarrow t = \sqrt[n]{0{,}95}$$

b) Simulamos $n = 20$ con $\theta = 3$ y $t = \dfrac{\text{máx}\{x_i\}_{i=1}^n}{2{,}5} = 1{,}17$

Calculamos $k_{0{,}05} = \sqrt[20]{0{,}95} = 0{,}9974$

Como $t = 1{,}17 > k_{0{,}05} = 0{,}9974 \Rightarrow$ se rechaza H_0



$$\text{p-valor} = P_{\theta=2{,}5}\big(T > \overbrace{1{,}17}^{t_m}\big) = 0$$

Ejercicio 4

De un experimento en los efectos de un medicamento para la ansiedad se midió el puntaje de un test de memoria antes y después de tomar el medicamento.

1. A partir de los datos que se encuentran en el archivo Islander_data.csv, diseñar un test de hipótesis de nivel de significación 0.01 para decidir si el tiempo medio de respuesta después de tomar el medicamento es menor que antes de tomarlo.
2. Hallar el p-valor

X_A: puntaje antes , X_E: puntaje después

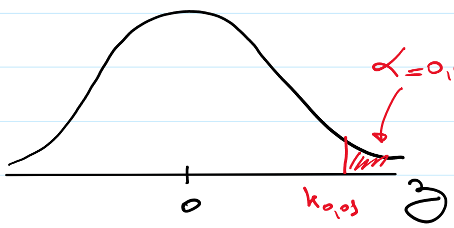
μ_A: tiempo medio de X_A , μ_E: tiempo medio de X_B

Defino una nueva variable: $X_D = X_E - X_A$

$H_0: \mu_D \leq 0$ vs. $H_1: \mu_D > 0$

Estadístico de prueba: $\frac{\overline{X}-\mu}{S} \cdot \sqrt{n} \overset{(a)}{\sim} N(0;1)$

Por T.C.L.



$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \frac{\overline{x}-0}{s}\sqrt{198} \geq k_{0,01} = \\ 0 & \text{si } \frac{\overline{x}-0}{s}\sqrt{198} < k_{0,01} = \end{cases}$$

• Cálculo del estadístico a partir de la muestra:

$\overline{x}$ de la muestra

desvío de la muestra (calculado con Python)

$$\frac{2,9545}{10,74}\sqrt{198} = 3,86 > k_{0,01} = 2,33 \Rightarrow \text{se rechaza } H_0$$

• p-valor $= P_{\mu_D=0}(Z > 3,86) \simeq 0$