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 0.95 para p
basado en la observación x=50.

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

Parámetro a estimar

$$\text{Estimador de } p: \quad \overline{X} = \frac{\sum_{i=1}^{50} X_i}{50} = \hat{p} \qquad \sum_{i=1}^{50} X_i \sim Bi(50;p)$$

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

Por Teorema Central del Límite

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

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

$$U = \frac{\overline{X} - p}{\sqrt{\underbrace{\hat{p}}_{\overline{X}}(1-\underbrace{\hat{p}}_{\overline{X}})}} \cdot \sqrt{50} \overset{(a)}{\sim} \mathcal{N}(0;1)$$

por teorema slutsky

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

- Se obtuvo de la muestra: $\overline{x} = \frac{50}{50} = 1$

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

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

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

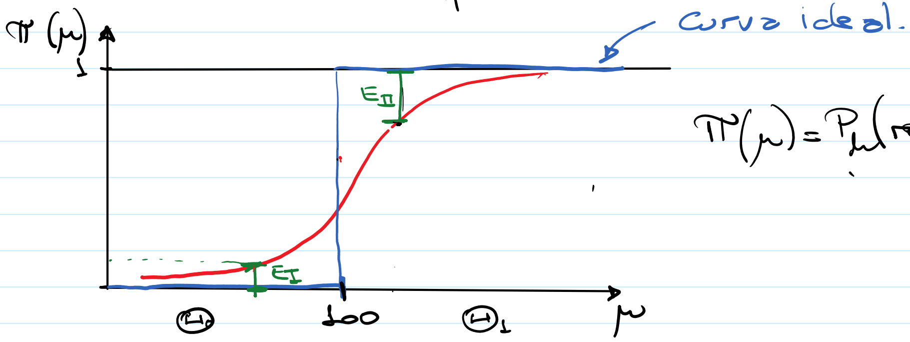
X_i: resistencia de la soldadura i (lb/plg), $i=1,2,\dots,n$

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

Error tipo I: decir que la planta pasa la prueba ($\mu > 100$) cuando en realidad la resistencia media no es mayor a 100

Error tipo II: decir la planta no pasa la prueba cuando la resistencia media es mayor a 100.

Curva de potencia



$\pi(\mu) = P_\mu(\text{rechazar } H_0)$

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.

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

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

Estimador de p: $\sum^{100} X_i$

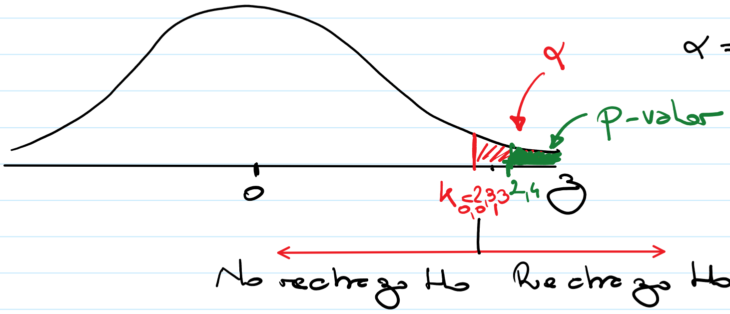
$H_0: p \leq 0{,}5$ vs $H_1: p > 0{,}5$

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

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

Por T.C.L.

$$\underbrace{\delta(\underline{X})}_{\text{Bajo Ho verdadera}} = \begin{cases} 1 & \text{si } \dfrac{\bar{X}-0{,}5}{\sqrt{0{,}5\cdot 0{,}5}}\cdot\sqrt{100} > k_\alpha = k_{0{,}01} = 2{,}33 \\ 0 & \text{si } \dfrac{\bar{X}-0{,}5}{\sqrt{0{,}5\cdot 0{,}5}}\cdot\sqrt{100} \leq k_\alpha = k_{0{,}01} = 2{,}33 \end{cases}$$



$$\alpha = P_{p=0{,}5}\left(\delta(\underline{X})=1\right) = 0{,}01$$

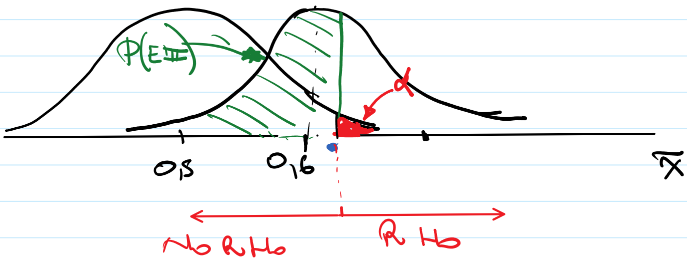
. De la muestra se obtuvo $\bar{X} = \hat{p} = 0{,}62$, entonces:

$$\frac{0{,}62-0{,}5}{\sqrt{0{,}5\cdot 0{,}5}}\sqrt{100} = 2{,}4 > 2{,}33 \Rightarrow \text{rechazamos Ho.}$$

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

$$= 0{,}0082$$

. Si $p = 0{,}6$, calculo $P(E_{II})$:



$$P(E_{II}) = P_{p=0{,}6}\left(\delta(\underline{X})=0\right) = P_{p=0{,}6}\left(\frac{\bar{X}-0{,}5}{0{,}5}\cdot 10 < 2{,}33\right) =$$

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

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

$$= \Phi(0{,}33) = 0{,}6293$$

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.

Parámetro bajo prueba.

$$X_1, X_2, \dots, X_n \overset{iid}{\sim} \mathcal{U}(0,\theta) \quad , i=1,2,\dots,n$$

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

Estimador de M.V. de θ: $\hat{\theta} = \max\{\underline{X}\} = U$

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

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

Propongo un estadístico de prueba: $T = \dfrac{U}{\theta}$

$$F_T(t) = P(T \leq t) = P\left(\frac{U}{\theta} \leq t\right) = P(U \leq t\cdot\theta) =$$

$$= F_U(t\cdot\theta) = \left(\frac{t\cdot\cancel{\theta}}{\cancel{\theta}}\right)^n \Rightarrow F_T(t) = t^n$$

$$\underbrace{\delta(\underline{X})}_{\text{Bajo } H_0 \text{ verdadero.}} = \begin{cases} 1 & \text{si } T = \dfrac{U}{2{,}5} > k_\alpha = \sqrt[n]{0{,}95} \\ 0 & \text{si } T = \dfrac{U}{2{,}5} \leq k_\alpha = \sqrt[n]{0{,}95} \end{cases}$$

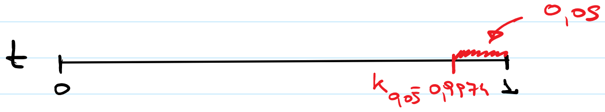
$$\alpha = P_{\theta=2{,}5}\left(\delta(\underline{X}) = 1\right) = P_{\theta=2{,}5}\left(\frac{U}{\theta} > k_\alpha\right) = 0{,}05 \Rightarrow$$

$$\Rightarrow P_{\theta=2{,}5}(T > k_\alpha) = 0{,}05 \Rightarrow 1 - F_T(t) = 0{,}05 \Rightarrow$$

$$\Rightarrow 1 - t^n = 0{,}05 \Rightarrow t = \sqrt[n]{0{,}95}$$

b. $\theta = 3 \rightarrow$ simulamos $n=20$ y obtenemos el valor del estadístico T.

$$t = \frac{\max(\underline{x})}{2{,}5} = 1{,}1584 > k_{0{,}05} = 0{,}9974 \Rightarrow \text{rechazamos } H_0.$$



$$p\text{-valor} = P_{\theta=2,5}(T > 1,1584) = 0$$

- Otro ejemplo: $n=20$ y se obtiene $\frac{\max\{\underline{x}\}}{2,5} = 0,9998$

$$p\text{-valor} = P(T > 0,9998) = 1 - F_T(0,9998) = 1 - 0,9998^{20} = 0,004$$

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_i: puntaje antes de la persona i. $\rightarrow \mu_X$ media de X

Y_i: " después de la persona i $\rightarrow \mu_Y$ media de Y

$H_0: \mu_Y \geq \mu_X$ vs. $H_1: \mu_Y < \mu_X$

$H_0: \underbrace{\mu_Y - \mu_X}_{D} \geq 0$ vs. $H_1: \underbrace{\mu_Y - \mu_X}_{D} < 0$

D: el valor medio de $Y_i - X_i$