p-valor para una prueba de hipótesis

Dada una prueba de hipótesis con región crítica

$$RC = \{ \underbrace{T(X_1, \dots, X_n)}_{\substack{\text{estadístico que} \\ \text{depende de} \\ \text{la muestra}}} \geq \underbrace{c}_{\substack{\text{no depende} \\ \text{de la muestra} \\ \text{(depende de } \alpha)}} \}$$

$(x_1, \dots, x_n)$ observaciones

$t_0 = T(x_1, \dots, x_n) \rightsquigarrow$ valor observado de T

$$\text{p-valor} = P_{H_0}(T \geq t_0)$$

* rechazo $H_0 \iff \alpha \geq$ p-valor
* no rechazo $H_0 \iff$ p-valor $\leq \alpha$

Ejercicio 6

Se debe reparar una máquina en una fábrica cuando produce más de 10 % de piezas defectuosas en un lote grande de artículos producido diariamente. Una muestra aleatoria de 100 artículos de la producción del día contiene 15 piezas defectuosas y el supervisor dice que se debe reparar la máquina. Contrastar la hipótesis nula "la proporción de piezas defectuosas es menor o igual al 10 %" contra la hipótesis alternativa "la proporción de piezas defectuosas es mayor que el 10 %".

$X \sim Ber(p)$ donde p es la proporción verdadera de piezas defectuosas

$$\begin{cases} H_0: & p \leq 0{,}10 \\ H_1: & p > 0{,}10 \end{cases} = \begin{cases} H_0: & p = 0{,}10 \\ H_1: & p > 0{,}10 \end{cases}$$

la región crítica es de la forma

$$\begin{cases} H_0 & : \quad p \leq p_0 \; 0{,}10 \\ H_1 & : \quad p > p_0 \end{cases} \qquad RC = \left\{ \overline{X}_n \geq p_0 + \frac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} z_\alpha \right\}$$

$$RC = \left\{ \overline{X}_n \geq 0{,}1 + \frac{\sqrt{0{,}1(1-0{,}1)}}{\sqrt{n}} z_\alpha \right\}$$

* vamos a calcular el p-valor

$$RC = \left\{ \overline{X_n} \geq \underbrace{0{,}1 + \frac{\sqrt{0{,}1(1-0{,}1)}}{\sqrt{n}} z_\alpha}_{\text{no depende de la muestra}} \right\}$$

$\uparrow$ depende de la muestra

tomamos $T(X_1, \ldots, X_n) = \overline{X_n}$

$t_0 = \overline{X_n} = \frac{15}{100} = 0{,}15$ ← proporción observada de piezas defectuosas

$X \sim Ber(p) \to E(X) = p$
$\to Var(X) = p(1-p)$

$$\text{p-valor} = P_{H_0}(T \geq t_0)$$

$$= P_{H_0}(\overline{X_n} \geq 0{,}15)$$

estamos suponiendo que H_0 es cierta

$X \sim Ber(0{,}1)$

$\overline{X_n} \sim \mathcal{N}\left(0{,}1 , \frac{0{,}1 \cdot 0{,}9}{100}\right)$

$\overline{X_n} \sim \mathcal{N}(0{,}1 , 0{,}0009)$

$$= P\Bigg(\underbrace{\frac{\overline{X_n} - 0{,}1}{\sqrt{0{,}0009}}}_{\mathcal{N}(0,1)} \geq \frac{0{,}15 - 0{,}1}{\sqrt{0{,}0009}}\Bigg)$$

$$= 1 - \phi\left(\frac{0{,}15 - 0{,}1}{\sqrt{0{,}0009}}\right)$$

$$= 1 - \phi(1{,}66)$$

$$= 1 - 0{,}9515$$

$$= 0{,}0485$$

$\boxed{\text{p-valor} = 0{,}0485}$

rechazo $H_0 \iff \alpha > \text{p-valor}$

no rechazo $H_0 \iff \text{p-valor} > \alpha$

* si $\alpha = 0{,}1$ tenemos $\alpha > \text{p-valor} \Rightarrow$ rechazamos H_0
* si $\alpha = 0{,}05$ tenemos $\alpha > \text{p-valor} \Rightarrow$ rechazamos H_0
* no rechazamos H_0 si $\alpha < 0{,}0485$

segundo parcial 2020

Ejercicio 7 (11 puntos)

En una fábrica embotelladora de agua mineral se utiliza una máquina para verter el líquido dentro de las botellas. La máquina vierte el líquido con un error aleatorio cuya distribución es normal de media cero y desvío conocido de 15 ml. En su funcionamiento correcto la máquina debe verter en promedio $\mu = 1.5$ litros de agua en cada botella.

Para un control rutinario de calidad se desea tomar una muestra de n botellas, con el objetivo de ralizar el siguiente test de hipótesis:

$$\begin{cases} H_0 : \mu = 1500 \text{ (ml)} \\ \boxed{H_A : \mu < 1500 \text{ (ml)}} \end{cases}$$

$$\begin{cases} H_0 : \mu = 1500 \\ H_1 : \mu = 1490 \end{cases}$$

Se decide utilizar una región de rechazo de la forma $\{\bar{X} \leq 1500 - \varepsilon\}$, en donde $\bar{X}$ es el promedio de las n botellas, y un nivel de significancia $\underline{\alpha = 5\%}$.

Suponga que la realidad es que la máquina está descalibrada y vierte en promedio $\underline{\mu = 1490}$ ml. Hallar el valor de n para que la potencia del test anterior sea de 55%.

(A) 8 **(B)** 13 **(C)** 17 **(D)** 25 **(E)** 36 **(F)** 15

X = cantidad de líquido en ml que la máquina vierte dentro de una botella.

$X = \mu + e_i \qquad e_i \sim \mathcal{N}(0, 15^2)$

$X \sim \mathcal{N}(\mu, 15^2)$

$$\begin{cases} H_0 : \mu = 1500 \\ H_1 : \mu < 1500 \end{cases}$$

tomamos $\underline{\alpha = 0{,}05}$

7.1.1 Pruebas sobre la media

1. Caso σ^2 conocido.

   - $$H_0 : \mu \leq \mu_0 \qquad H_1 : \mu > \mu_0 \qquad RC = \left\{ \overline{X}_n \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$$

   - $$H_0 : \mu \geq \mu_0 \qquad H_1 : \mu < \underbrace{\mu_0}_{1500} \qquad RC = \left\{ \overline{X}_n \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha \right\}$$

entonces la región crítica es de la forma

$$RC = \left\{ \overline{X_n} \leq 1500 - \underbrace{\frac{15}{\sqrt{n}} z_\alpha}_{\varepsilon} \right\}$$

tenemos $\alpha = 0{,}05$

$$z_\alpha = \phi^{-1}(1-\alpha) = \phi^{-1}(0{,}95) = 1{,}645 \qquad \varepsilon = \frac{15}{\sqrt{n}} \cdot 1{,}645$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.0
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9

en realidad $\mu = 1490$

buscamos n para que la prueba tenga potencia $0{,}55$

potencia = P(rechazar H_0 si H_1 es cierta)

$$= P_{H_1}(\text{rechazar } H_0)$$

$$= P_{H_1}\left(\overline{X_n} \leq 1500 - \varepsilon \right)$$

sabemos que $\mu = 1490$

$X \sim \mathcal{N}(1490, 15^2)$

$\overline{X_n} \sim \mathcal{N}\left(1490, \frac{15^2}{n}\right)$

$$= P_{H_1}\left(\frac{\overline{X_n} - 1490}{15/\sqrt{n}} \leq \frac{1500 - \varepsilon - 1490}{15/\sqrt{n}} \right)$$

$$= P_{H_1}\left(\frac{\overline{X_n} - 1490}{15}\sqrt{n} \leq \frac{10 - \varepsilon}{15}\sqrt{n} \right)$$

$$= \phi\left(\frac{10 - \varepsilon}{15}\sqrt{n} \right)$$

entonces

$$0,55 = \phi\left(\frac{10-\varepsilon}{15}\sqrt{n}\right)$$

$$\phi^{-1}(0,55) = \frac{10-\varepsilon}{15}\sqrt{n}$$

$$0,125 = \frac{10-\varepsilon}{15}\sqrt{n}$$

$$0,125 = \frac{10-\varepsilon}{15}\sqrt{n} \Rightarrow 10-\varepsilon = 0,125 \cdot \frac{15}{\sqrt{n}}$$

$$\Rightarrow 10 - \frac{15}{\sqrt{n}} \cdot 1,645 = 0,125 \cdot \frac{15}{\sqrt{n}}$$

$$\Rightarrow 10 = 0,125 \cdot \frac{15}{\sqrt{n}} + \frac{15}{\sqrt{n}} \cdot 1,645$$

$$\Rightarrow 10 = \frac{15}{\sqrt{n}}(0,125 + 1,645)$$

$$\Rightarrow \frac{1}{\sqrt{n}} = \frac{10}{15(0,125+1,645)}$$

$$\Rightarrow n = \left(\frac{15(0,125+1,645)}{10}\right)^2 = 7,05$$

entonces el menor tamaño de muestra es $\underline{\underline{n=8}}$

Ejercicio 8

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

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

$X \sim \mathcal{U}(0,b)$

$X_1, \ldots, X_{200}$ MAS de X

a) $\overline{X_n}$ es un estimador para $m = E(X)$

vamos a buscar un intervalo de confianza centrado en $\overline{X_n}$

IdC: $[\overline{X_n}-k \, , \, \overline{X_n}+k]$

$$P(m \in [\overline{X_n}-k, \overline{X_n}+k]) = 1-\alpha$$

$$1-\alpha = P(m \in [\overline{X_n}-k, \overline{X_n}+k])$$

$$= P(\overline{X_n}-k \leq m \leq \overline{X_n}+k)$$

$$= P(-k \leq m-\overline{X_n} \leq k)$$

$$= P(-k \leq \overline{X_n}-m \leq k)$$

$$= P\Big(\frac{-k}{\sigma}\sqrt{n} \leq \underbrace{\frac{\overline{X_n}-m}{\sigma}\sqrt{n}}_{\sim \mathcal{N}(0,1)} \leq \frac{k}{\sigma}\sqrt{n}\Big)$$

$X \sim \mathcal{U}(0,b)$ $\to m = E(X) = \frac{b}{2}$, $\to Var(X) = \frac{b^2}{12} = \sigma^2$

$\overline{X_n} \sim \mathcal{N}(m, \frac{\sigma^2}{n})$

$$= \phi\Big(\frac{k}{\sigma}\sqrt{n}\Big) - \phi\Big(-\frac{k}{\sigma}\sqrt{n}\Big)$$

$$= \phi\Big(\frac{k}{\sigma}\sqrt{n}\Big) - \Big(1-\phi\Big(\frac{k}{\sigma}\sqrt{n}\Big)\Big)$$

$$= 2\phi\Big(\frac{k}{\sigma}\sqrt{n}\Big) - 1$$

entonces: $1-\alpha = 2\phi\Big(\frac{k}{\sigma}\sqrt{n}\Big) - 1$

$$\Rightarrow 2-\alpha = 2\phi\Big(\frac{k}{\sigma}\sqrt{n}\Big)$$

$$\Rightarrow 1-\frac{\alpha}{2} = \phi\Big(\frac{k}{\sigma}\sqrt{n}\Big)$$

$$\Rightarrow \underbrace{\phi^{-1}\Big(1-\frac{\alpha}{2}\Big)}_{z_{\alpha/2}} = \frac{k}{\sigma}\sqrt{n}$$

$$\Rightarrow k = \frac{z_{\alpha/2}\cdot\sigma}{\sqrt{n}}$$

tomamos $k = \dfrac{z_{\alpha/2}\, S_n}{\sqrt{n}}$