

Número	Nombre	CI	Nota
--------	--------	----	------

SEGUNDO PARCIAL - 26 de Noviembre de 2022

Ejercicio 1 Para un conjunto de datos se tiene la siguiente información:

- Cuantil 0.75 = 6.3
- Cuantil 0.25 = 4.3

- (a) Calcular el rango intercuantil.
- (b) ¿Cuales de los siguientes datos {8,1; 6,9; 12; 3,3; 0,9} son considerados como atípicos (outliers) a partir de un boxplot (para un conjunto de datos con los cuantiles anteriores)?
- (c) ¿Cambiarían los atípicos si el cuantil 0.75 fuera 7.3 y el cuantil 0.25 fuera 5.3? Indique cuáles serían los datos atípicos en este caso.

Ejercicio 2¹ Para analizar si el fumar disminuye el peso de los bebés al nacer se observó un grupo formado por madres que habían fumado por lo menos dos años hasta el momento de quedar embarazadas. Para el caso de bebés de madres no fumadoras se conoce la media $\mu_0 = 2215$ y el desvío estandar $\sigma_0 = 96$ del peso al nacer (en gramos).

Se cuenta con una muestra de tamaño 20 de madres fumadoras para la cual se calcularon la media \bar{x}_n y el desvío estandar s_n a partir de los datos.

	Fumadoras	No fumadoras
Media	$\bar{x}_n = 1910$	$\mu_0 = 2215$
desvío	$s_n = 134$	$\sigma_0 = 96$

Sea μ la media de los pesos al nacer de los bebés de madres fumadoras.

- (a) Hallar un intervalo de confianza de nivel 0,95 para μ .
- (b) Testear al nivel $\alpha = 0,05$ las hipótesis $H_0 : \mu = 2215$ vs $H_A : \mu < 2215$.

Ejercicio 3 De una determinada empresa se conocen los siguientes datos: volumen de ventas y gasto en publicidad (en millones de pesos) de los últimos 6 años. Se quiere analizar el volumen de ventas obtenido en función del gasto invertido en publicidad.

Ventas	Publicidad
10	16
15	32
20	48
22	56
30	64
32	80

- (a) Indicar cuál sería la variable explicativa (X) y cuál la variable explicada (Y) en un modelo de regresión lineal para estos datos.
- (b) Calcular los estimadores \hat{a} y \hat{b} y hallar la recta de regresión lineal.
- (c) ¿Qué volumen de ventas de la empresa se podría esperar para un año en donde se gasten 60 millones en publicidad? ¿Y para un gasto en publicidad de 100 millones?
- (d) ¿A cuál de las dos predicciones le tiene más confianza? Justifique.

¹Los siguientes datos pueden ser útiles para este ejercicio: $qnorm(0.95)=1.65$, $qnorm(0.975)=1.96$, $qnorm(0.90)=1.28$.

SOLUCIÓN

Ejercicio 1

(a) El rango intercuartílico es $RIC = Q_{0,75} - Q_{0,25} = 6,3 - 4,3 = 2$.

(b) En el boxplot:

$$L_{inf} = Q_{0,25} - 1,5 \times RIC = 4,3 - 1,5 \times 2 = 1,3 \text{ y } L_{sup} = Q_{7,25} + 1,5 \times RIC = 6,3 + 1,5 \times 2 = 9,3$$

Los *datos atípicos* son los datos x_i tales que $x_i < L_{inf}$ ó $x_i > L_{sup}$. Luego, 12 y 0,9 son datos atípicos.

(c) Si fuera $Q_{0,75} = 7,5$ y $Q_{0,25} = 5,3$, entonces $RIC = 2,2$, $L_{inf} = 2$, $L_{sup} = 10,8$ y los datos atípicos no cambian.

Ejercicio 2

(a) Un intervalo de confianza aproximado a nivel 0,95 para μ es:

$$I_{0,95} = [\bar{x}_n \pm k] \text{ con } k = \frac{s_n}{\sqrt{n}} z_{1-\frac{\alpha}{2}} = \frac{134}{\sqrt{20}} 1,96 \approx 59.$$

Luego, $I_{0,95} \approx [1910 \pm 59] = [1851, 1969]$.

(b) Queremos testear $\begin{cases} H_0 : \mu = 2215 \\ H_A : \mu < 2215 \end{cases}$. Hay dos formas de resolver este ejercicio.

- *Mediante la construcción de una región crítica*, que en este caso será

$$RC = \{(x_1, \dots, x_n) : \bar{x}_n < k\} \text{ con } k \text{ tal que } \mathbb{P}(RC|H_0) = \alpha,$$

y esto se cumple si

$$k = \mu_0 - \frac{z_{1-\alpha}}{\sqrt{n}} s_n = 2215 - \frac{qnorm(0,95)}{\sqrt{20}} \times 134 = 2215 - \frac{1,65}{\sqrt{20}} \times 134 = 2166$$

Luego $RC = \{(x_1, \dots, x_n) : \bar{x}_n < 2166\}$. Como $\bar{x}_n = 1910$, entonces la muestra pertenece a RC y rechazamos H_0 .

- *Mediante el cálculo del p-valor*: En este caso:

$$\begin{aligned} p\text{-valor} &= \mathbb{P}(\bar{X}_n \leq 1910|H_0) = \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \leq \sqrt{n}\frac{1910 - \mu}{\sigma}|H_0\right) \\ &= \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} \leq \sqrt{n}\frac{1910 - \mu_0}{\sigma}|H_0\right) \approx \Phi\left(\sqrt{n}\frac{1910 - \mu_0}{\sigma}\right) \\ &\approx \Phi\left(\sqrt{n}\frac{1910 - \mu_0}{s_n}\right) = \Phi\left(\sqrt{20}\frac{1910 - 2215}{134}\right) = \Phi(-10,2) \approx 0. \end{aligned}$$

Como el p-valor es casi cero, entonces $\alpha > p\text{-valor}$ y rechazamos H_0 .

Ejercicio 3

- (a) Son X =Publicidad e Y =Ventas, puesto que se pretende estudiar si la publicidad afecta efectivamente a las ventas.
- (b) Asumimos un modelo de la forma $Y = aX + b + \varepsilon$. Los estimadores de a y b son:

$$\hat{a}_n = \frac{\sum_{i=1}^n x_i y_i - n \bar{x}_n \bar{y}_n}{\sum_{i=1}^n x_i^2 - n(\bar{x}_n)^2} \quad \hat{b}_n = \bar{y}_n - \hat{a}_n \bar{x}_n$$

En este caso:

- $n = 6$
- $\bar{x}_n = 49,3$
- $\bar{y}_n = 21,5$
- $\sum x_i y_i = 7312$
- $\sum x_i^2 = 17216$

Luego, $\hat{a}_n \approx 0,4$ y $\hat{b}_n \approx 1,8$ y el modelo es $y = r(x) = 0,4x + 1,8$.

- (c) $r(60) = 25,8$ y $r(100) = 41,8$.
- (d) Genera más confianza $r(60)$ puesto que tenemos datos cercanos a 60 (por arriba y abajo), pero no de 100.