Intervalos y Regiones de Confianza

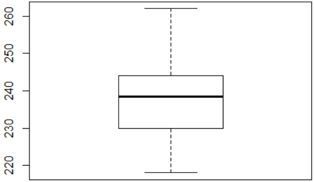
Ejemplo de duración de baterías

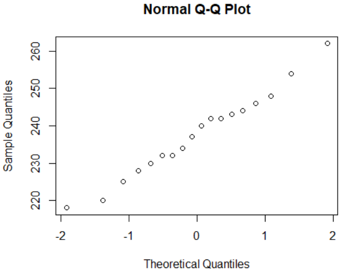
- Los siguientes datos corresponden a la duración en horas de 18 baterías eléctricas (Maronna, 2021):

  237 242 232 242 248 230 244 243 254
  262 234 220 225 246 232 218 228 240

- Queremos estimar la duración media.

Ejemplo: continuación





Ejemplo: continuación

- Modelo: $X_i \sim \mathcal{N}(\mu, \sigma^2)$
- Asumimos por ahora: $\sigma = \sigma_o = 10$
- Estimación con los datos: $\bar{x}_n = 237.6111$
- ¿μ vale 237.6111?
- **Objetivo:** pasar de la **estimación puntual** a la **estimación por intervalo**.
- Daremos un **intervalo de valores compatibles con** μ.

Intervalos de confianza: definición

Sea $\mathbf{X}_n = \{X_1, \ldots, X_n\}$ una muestra aleatoria.
Diremos que $(a(\mathbf{X}_n), b(\mathbf{X}_n))$ es un intervalo de confianza de nivel $1-\alpha$ para el parámetro θ sii

$$\mathbb{P}\left(a(\mathbf{X}_n) \ < \ \theta \ < \ b(\mathbf{X}_n)\right) = 1 - \alpha \ .$$

Intervalos de confianza: definición

Sea $\mathbf{X}_n = \{X_1, \ldots, X_n\}$ una muestra aleatoria.
Diremos que $(a(\mathbf{X}_n), b(\mathbf{X}_n))$ es un intervalo de confianza de nivel $1-\alpha$ para el parámetro θ sii

$$\mathbb{P}\left(a(\mathbf{X}_n) \ < \ \theta \ < \ b(\mathbf{X}_n)\right) = 1 - \alpha \ .$$

La longitud de $\mathcal{S}(\mathbf{X}_n) = (a(\mathbf{X}_n), b(\mathbf{X}_n))$ es

$$l = b(\mathbf{X}_n) - a(\mathbf{X}_n)$$

Bajo Normalidad: σ^2 conocida

$\sigma^2 = \sigma_0^2$

- Buscamos intervalo de confianza para μ
- Pivote: función de la muestra y del parámetro de interés, cuya distribución es conocida.

$$\frac{\bar{X}_n - \mu}{\sqrt{\sigma_0^2/n}}\ , \quad \widehat{\mu}_n = \bar{X}_n$$

- Distribución del pivote:

$$\frac{\bar{X}_n - \mu}{\sqrt{\sigma_0^2/n}} \sim N(0,1)$$

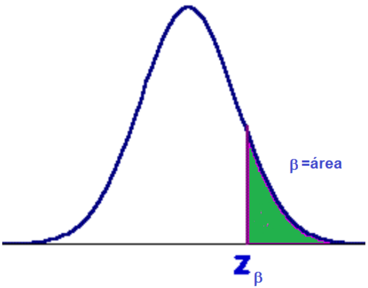
- Sea z_β con $P(Z > z_\beta) = \beta$: (veamos el gráfico)

$$z_\beta = \phi^{-1}(1-\beta) = \texttt{qnorm(1-beta)}$$

- Luego, tenemos que

$$\mathbb{P}\left(-z_{\alpha/2}\ <\ \frac{\bar{X}_n - \mu}{\sqrt{\sigma_0^2/n}}\ <\ z_{\alpha/2}\right) = 1-\alpha$$

Area bajo la $N(0, 1)$



Bajo Normalidad: σ^2 conocida

Por lo tanto, tenemos que

$$\mathbb{P}\left(\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\sigma_0^2}{n}} \ < \ \mu \ < \ \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\sigma_0^2}{n}}\right) = 1-\alpha$$

Así, obtenemos que

$$\left(\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\sigma_0^2}{n}} \quad , \quad \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\sigma_0^2}{n}}\right)$$

es un intervalo de confianza de nivel $1-\alpha$ para μ bajo el modelo normal cuando la varianza es σ_0^2 conocida.

Veamos mediante resultados de una simulación cómo se interpreta la confianza y el cubrimiento empírico

Bajo Normalidad: σ^2 conocida

Intervalo de confianza de nivel $1-\alpha$ para μ

$$\left(\overline{X}_n - z_{\alpha/2}\sqrt{\frac{\sigma_0^2}{n}} \quad , \quad \overline{X}_n + z_{\alpha/2}\sqrt{\frac{\sigma_0^2}{n}} \right)$$

Algunas preguntas

- ¿Cuánto vale longitud tiene el intervalo? $l =$
- ¿Qué ocurre con l a medida que n aumenta?
- ¿Qué ocurre con l cuando el nivel $1-\alpha$ aumenta?
- ¿Qué ocurre con l cuando aumenta la varianza σ_0^2?

Bajo Normalidad: σ^2 conocida

¿Y si queremos que la longitud sea menor a cierto valor l_o?

- $X_i \sim \mathcal{N}(\mu, \sigma_o^2)$ i.i.d. Buscamos intervalo de confianza para μ.
- σ_o conocido: IC nivel $1-\alpha$ para μ

$$\left(\bar{X}_n - z_{\alpha/2}\,\frac{\sigma_o}{\sqrt{n}} \quad , \quad \bar{X}_n + z_{\alpha/2}\,\frac{\sigma_o}{\sqrt{n}}\right)$$

- longitud $\rightarrow l = 2\, z_{\alpha/2}\, \frac{\sigma_o}{\sqrt{n}}$

Bajo Normalidad: σ^2 conocida

¿Y si queremos que la longitud sea menor a cierto valor l_o?

- $X_i \sim \mathcal{N}(\mu, \sigma_o^2)$ i.i.d. Buscamos intervalo de confianza para μ.
- σ_o conocido: IC nivel $1-\alpha$ para μ

$$\left(\bar{X}_n - z_{\alpha/2}\, \frac{\sigma_o}{\sqrt{n}} \quad , \quad \bar{X}_n + z_{\alpha/2}\, \frac{\sigma_o}{\sqrt{n}} \right)$$

- longitud $\rightarrow l = 2\, z_{\alpha/2}\, \frac{\sigma_o}{\sqrt{n}} \leq l_o$

Bajo Normalidad: σ^2 conocida

¿Y si queremos que la longitud sea menor a cierto valor l_o?

- $X_i \sim \mathcal{N}(\mu, \sigma_o^2)$ i.i.d. Buscamos intervalo de confianza para μ.
- σ_o conocido: IC nivel $1-\alpha$ para μ

$$\left(\bar{X}_n - z_{\alpha/2}\,\frac{\sigma_o}{\sqrt{n}} \quad , \quad \bar{X}_n + z_{\alpha/2}\,\frac{\sigma_o}{\sqrt{n}}\right)$$

- longitud $\rightarrow l = 2\, z_{\alpha/2}\, \frac{\sigma_o}{\sqrt{n}} \leq l_o$
- Si queremos $l \leq l_o \rightarrow \frac{4 z_{\alpha/2}^2 \sigma_o^2}{l_o^2} \leq n$

Regiones de confianza

Dado un vector $\mathbf{X}$ con distribución perteneciente a la familia $F(\mathbf{x}, \theta)$ con $\theta \in \Theta$, *una región de confianza* $\mathcal{S}(\mathbf{X})$ *para* θ *con nivel de confianza* $1-\alpha$ será una función que a cada $\mathbf{X}$ le asigne un subconjunto $\mathcal{S}(\mathbf{X}) \subset \Theta$ de manera que

$$\mathbb{P}_\theta(\theta \in \mathcal{S}(\mathbf{X})) = 1-\alpha, \qquad \forall \theta \in \Theta$$

Es decir, $\mathcal{S}(\mathbf{X})$ cubre el valor verdadero del parámetro con probabilidad $1-\alpha$.

Caso particular: Si $\theta \in \mathbb{R}$ se dirá que $\mathcal{S}(\mathbf{X})$ es un intervalo de confianza cuando

$$\mathcal{S}(\mathbf{X}) = [a(\mathbf{X}), b(\mathbf{X})]$$

Procedimiento general: Método del Pivote

Sea $\mathbf{X}$ un vector aleatorio cuya distribución pertenece a la familia $F(\mathbf{x}, \theta)$, $\theta \in \Theta$.
Una función $G(\mathbf{X}, \theta)$ se llama un pivote sii la distribución de $G(\mathbf{X}, \theta)$ no depende de θ.

Procedimiento general: Método del Pivote

Sea $\mathbf{X}$ un vector aleatorio cuya distribución pertenece a la familia $F(\mathbf{x}, \theta)$, $\theta \in \Theta$.
Una función $G(\mathbf{X}, \theta)$ se llama un pivote sii la distribución de $G(\mathbf{X}, \theta)$ no depende de θ.

Teorema *Sea* $\mathbf{X}$ *un vector aleatorio cuya distribución pertenece a la familia* $F(\mathbf{x}, \theta)$, $\theta \in \Theta$. *Sea*

- $U = G(\mathbf{X}, \theta)$ *una variable aleatoria cuya distribución es independiente de* θ.
- A *y* B *tales que* $\mathbb{P}(A \leq U \leq B) = 1 - \alpha$.

Luego, si $\mathcal{S}(\mathbf{X}) = \{\theta : A \leq G(\mathbf{X}, \theta) \leq B\}$, $\mathcal{S}(\mathbf{X})$ *es una región de confianza a nivel* $(1 - \alpha)$ *para* θ.

¿Qué hacemos si σ es desconocida?

Si no conocemos la varianza σ^2, deberemos reemplazar a σ^2 por un estimador....

Usemos $S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{n-1}$, pero...

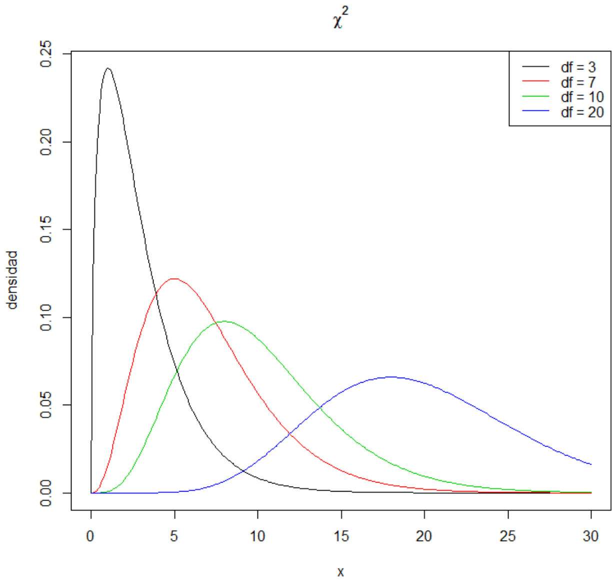
$$\frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \sim \quad ?$$

Veamos los resultados de una simulación

Distribución χ_k^2

- Sean $Z_1, \ldots, Z_k$ i.i.d., $Z_i \sim N(0,1)$.
- Llamamos χ_k^2 (chi cuadrado con $k-$grados de libertad) a la distribución de
$$Z_1^2 + Z_2^2 + \cdots + Z_k^2$$
o sea:
$$Z_1^2 + Z_2^2 + \cdots + Z_k^2 \sim \chi_k^2 \ .$$
- χ_k^2 corresponde a una $\Gamma(\frac{k}{2}, \frac{1}{2})$

χ^2_k



χ²
densidad
x
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25
30
df = 3
df = 7
df = 10
df = 20

Distribución t

Una variable aleatoria tiene ditribución t_k de Student con k grados de libertad si para todo $u \in \mathbb{R}$ su densidad es de la forma

$$f(u, k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{u^2}{k}\right)^{-\frac{k+1}{2}}$$

Distribución t

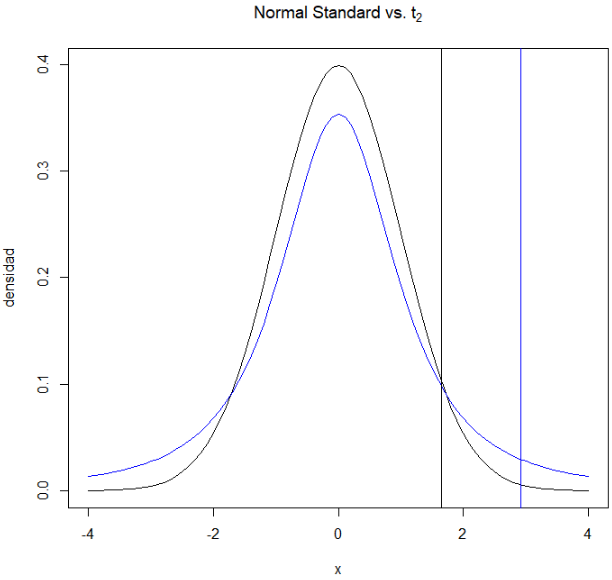
Una variable aleatoria tiene ditribución t_k de Student con k grados de libertad si para todo $u \in \mathbb{R}$ su densidad es de la forma

$$f(u,k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{u^2}{k}\right)^{-\frac{k+1}{2}}$$

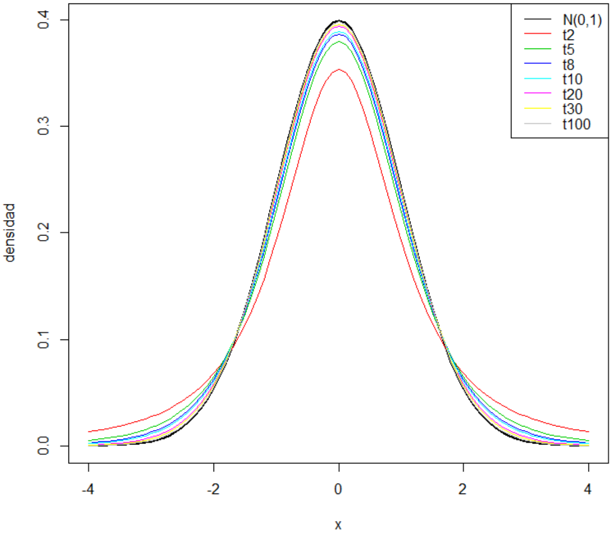
¿Cuándo aparece una t_k?
$Z \sim N(0,1)$ y $U \sim \chi^2_k$ independientes:

$$\frac{Z}{\sqrt{U/k}} \sim t_k$$

Mirando percentiles...



Comparando con la normal...



Intervalos para la media μ y la varianza de una normal

Buscando un pivote....

Intervalos para la media μ y la varianza de una normal

Buscando un pivote....

Teorema *Sean* $X_1, \ldots, X_n$ *variables aleatorias independientes donde* $X_i \sim N(\mu, \sigma^2)$*. Luego*

a) $Z = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1)$, *con* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

b) $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ *es independiente de* $\bar{X}_n$.

c) $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

d) $\sqrt{n}\ \frac{(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}$.

Recordemos que:

- Si Z es una v.a. con distribución normal standard, $Z^2 \sim \Gamma(1/2, 1/2)$, o sea χ^2_1.
- Si $U_1, \ldots, U_k$ independientes., $U_i \sim \Gamma(\alpha_i, \lambda)$, entonces

$$U_1 + U_2 + \cdots + U_k \sim \Gamma\left(\sum_{i=1}^{k} \alpha_i, \lambda\right)$$

- La suma de n v.a. independientes χ^2_1 tiene distribución χ^2_n.
- Si $Z_1, \ldots, Z_n$ i.i.d., $Z_i \sim \mathcal{N}(0, 1)$, entonces

$$Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \Gamma(n/2, 1/2) = \chi^2_n$$

Recordemos que:

- Si Z es una v.a. con distribución normal standard, $Z^2 \sim \Gamma(1/2, 1/2)$, o sea χ^2_1.
- Si $U_1, \ldots, U_k$ independientes., $U_i \sim \Gamma(\alpha_i, \lambda)$, entonces

$$U_1 + U_2 + \cdots + U_k \sim \Gamma\left(\sum_{i=1}^{k} \alpha_i, \lambda\right)$$

- La suma de n v.a. independientes χ^2_1 tiene distribución χ^2_n.
- Si $Z_1, \ldots, Z_n$ i.i.d., $Z_i \sim \mathcal{N}(0, 1)$, entonces

$$Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \Gamma(n/2, 1/2) = \chi^2_n$$

Consideremos las siguientes v.a. $Y_i = \frac{\bar{X}_i - \mu}{\sigma_o} \sim N(0, 1)$,

$$V = \sqrt{n}\bar{Y}, \quad W = \sum_{i=1}^{n} (Y_i - \bar{Y})^2, \quad U = \frac{V}{\sqrt{W/(n-1)}}$$

Bajo Normalidad con σ^2 desconocida

Buscamos un intervalo de confianza para μ con σ^2 desconocida.

- Pivote:

$$\frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \sim t_{n-1}$$

- Sea $t_{k,\beta}$ tal que $\mathbb{P}(Y > t_{k,\beta}) = \beta$ cuando $Y \sim t_k$, entonces

$$\mathbb{P}\left(-t_{n-1,\alpha/2} \ < \ \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \ < \ t_{n-1,\alpha/2}\right) = 1 - \alpha$$

Luego,

$$\left(\bar{X}_n - t_{n-1,\alpha/2}\sqrt{\frac{S_n^2}{n}} \quad , \quad \bar{X}_n + t_{n-1,\alpha/2}\sqrt{\frac{S_n^2}{n}}\right)$$

es un intervalo de confianza de nivel $1 - \alpha$ para μ bajo el modelo normal, con varianza σ^2 desconocida.

Veamos los resultados de una simulación