Intervalos de Confianza para la diferencia de medias de dos poblaciones normales

Ejemplo: Comparamos dos tratamientos

- **Objetivo**: comparar el rendimiento de dos fertilizantes.
- Se divide un terreno en 30 parcelas homogéneas por ej. cant. sol, humedad, etc.).
- En las 30 parcelas se cultiva la misma variedad de maíz.
- Eligiendo al azar, en 15 de ellas se utiliza el fertilizante A y en las restantes el B.

Ejemplo: Comparamos dos tratamientos

Datos

- Parcelas con Tratamiento A

$$238, 237, 235, 220, 233, 203, 228, 220, 221,$$
$$215, 218, 217, 232, 225, 209$$

- Parcelas con Tratamiento B

$$253, 227, 241, 245, 237, 248, 250, 218, 239,$$
$$243, 257, 208, 215, 240, 229$$

- Queremos comparar los dos fertilizantes mediante un intervalo de confianza para la diferencia de los rendimientos medios.

Modelo 1: Dos muestras normales independientes

Supongamos que:

- $X_1, \ldots, X_{n_1}, \quad X_i \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right)$
- $Y_1, \ldots, Y_{n_2}, \quad Y_i \sim \mathcal{N}\left(\mu_2, \sigma_2^2\right)$

son dos muestras aleatorias, independientes entre sí.

Intervalos para $\Delta = \mu_1 - \mu_2$

Caso 1 - varianzas conocidas σ_1^2 y σ_2^2 (iguales o no)

$$\overline{X}_{n_1} \sim \mathcal{N}\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \;, \quad \overline{Y}_{n_2} \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2^2}{n_2}\right) \text{ independientes,}$$

$$\overline{X}_{n_1} - \overline{Y}_{n_2} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\text{Pivote:} \quad \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0,1) \;,$$

Intervalo de nivel $1-\alpha$ para $\Delta = \mu_1 - \mu_2$:

$$\left(\overline{X}_{n_1} - \overline{Y}_{n_2} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \,,\, \overline{X}_{n_1} - \overline{Y}_{n_2} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Intervalos para $\Delta = \mu_1 - \mu_2$

Caso 2 - varianzas DESconocidas pero IGUALES
$\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}\left(0,1\right) \ , \quad \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}\left(0,1\right)$$

Necesitamos estimar σ^2.

Intervalos para $\Delta = \mu_1 - \mu_2$

Caso 2 - varianzas DESconocidas pero IGUALES
$\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0,1) \ , \quad \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}(0,1)$$

Necesitamos estimar σ^2. Tenemos dos estimadores:

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \overline{X}_{n_1})^2 \ , \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \overline{Y}_{n_2})^2$$

Intervalos para $\Delta = \mu_1 - \mu_2$

Caso 2 - varianzas DESconocidas pero IGUALES
$\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}\left(0,1\right) \;, \quad \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}\left(0,1\right)$$

Necesitamos estimar σ^2. Tenemos dos estimadores:

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X}_{n_1})^2 \;, \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y}_{n_2})^2$$

Los combinamos:

$$S_p^2 = \frac{S_1^2\left(n_1 - 1\right) + S_2^2\left(n_2 - 1\right)}{n_1 + n_2 - 2}$$

Estimador insesgado de σ^2

Ejercicio

Proposición: Sean $X_1, \ldots, X_{n_1}$ y $Y_1, \ldots, Y_{n_2}$ dos muestras aleatorias independientes de las distribuciones $N\left(\mu_1, \sigma^2\right)$ y $N\left(\mu_2, \sigma^2\right)$ respectivamente. Entonces,

$$S_p^2 = \frac{S_1^2\left(n_1 - 1\right) + S_2^2\left(n_2 - 1\right)}{n_1 + n_2 - 2}$$

es un estimador insesgado de σ^2.

Intervalos para $\Delta = \mu_1 - \mu_2$. El pivote

Caso 2: varianzas DESconocidas pero IGUALES: $\sigma^2 = \sigma_1^2 = \sigma_2^2$
Tenemos las siguientes variables independientes

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1), \quad \frac{S_1^2\,(n_1 - 1)}{\sigma^2} \sim \chi^2_{n_1 - 1}, \quad \frac{S_2^2\,(n_2 - 1)}{\sigma^2} \sim \chi^2_{n_2 - 1}$$

Intervalos para $\Delta = \mu_1 - \mu_2$. El pivote

Caso 2: varianzas DESconocidas pero IGUALES: $\sigma^2 = \sigma_1^2 = \sigma_2^2$
Tenemos las siguientes variables independientes

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1), \quad \frac{S_1^2\,(n_1-1)}{\sigma^2} \sim \chi^2_{n_1-1}, \quad \frac{S_2^2\,(n_2-1)}{\sigma^2} \sim \chi^2_{n_2-1}$$

Recordemos que suma de χ^2 independientes es χ^2. ¿Con cuántos grados de libertad?

Intervalos para $\Delta = \mu_1 - \mu_2$. El pivote

Caso 2: varianzas DESconocidas pero IGUALES: $\sigma^2 = \sigma_1^2 = \sigma_2^2$
Tenemos las siguientes variables independientes

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1), \quad \frac{S_1^2\,(n_1-1)}{\sigma^2} \sim \chi^2_{n_1-1}, \quad \frac{S_2^2\,(n_2-1)}{\sigma^2} \sim \chi^2_{n_2-1}$$

Recordemos que suma de χ^2 independientes es χ^2. ¿Con cuántos grados de libertad?

$$\frac{S_1^2\,(n_1-1) + S_2^2\,(n_2-1)}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

Recordemos que...

$Z \sim N(0,1)$ independiente de $U \sim \chi^2_k$

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

dicho en palabras

$$\frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2_k}{k}}} \sim t_k$$

cuando numerador y denominador son independientes.

Intervalos para $\Delta = \mu_1 - \mu_2$. El pivote

$$\frac{\dfrac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\sigma^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}}{\sqrt{\dfrac{S_1^2\,(n_1-1) + S_2^2\,(n_2-1)}{\sigma^2(n_1+n_2-2)}}} \sim t_{n_1+n_2-2}$$

Intervalos para $\Delta = \mu_1 - \mu_2$. El pivote

$$\frac{\dfrac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{\sigma^2 \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}}}{\sqrt{\dfrac{S_1^2 \left(n_1 - 1\right) + S_2^2 \left(n_2 - 1\right)}{\sigma^2 (n_1 + n_2 - 2)}}} \sim t_{n_1 + n_2 - 2}$$

Luego, si

$$S_p^2 = \frac{S_1^2 \left(n_1 - 1\right) + S_2^2 \left(n_2 - 1\right)}{n_1 + n_2 - 2}$$

tenemos que

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t_{n_1 + n_2 - 2}$$

Intervalos para $\Delta = \mu_1 - \mu_2$. El pivote.

Caso 2: varianzas desconocidas pero IGUALES: $\sigma^2 = \sigma_1^2 = \sigma_2^2$
PIVOTE:

$$\frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}\ , \quad S_p^2 = \frac{S_1^2\,(n_1 - 1) + S_2^2\,(n_2 - 1)}{n_1 + n_2 - 2}$$

Teorema

Sean $X_1, \ldots, X_{n_1}$ y $Y_1, \ldots, Y_{n_2}$ dos muestras aleatorias independientes de las distribuciones $N\left(\mu_1, \sigma^2\right)$ y $N\left(\mu_2, \sigma^2\right)$, respectivamente. Sean

$$U = \frac{\overline{X} - \overline{Y} - \Delta}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \quad V = \frac{\sum_{i=1}^{n_1} \left(X_i - \bar{X}\right)^2}{\sigma^2} \quad W = \frac{\sum_{i=1}^{n_2} \left(Y_i - \bar{Y}\right)^2}{\sigma^2}$$

$$S_p^2 = \frac{S_1^2\left(n_1 - 1\right) + S_2^2\left(n_2 - 1\right)}{n_1 + n_2 - 2}$$

Luego

(i) U, V y W son variables aleatorias independientes con distribuciones $N(0,1), \chi^2_{n_1-1}$ y $\chi^2_{n_2-1}$ respectivamente.

(ii) $Z = V + W \sim \chi^2_{n_1+n_2-2}$.

(iii) $\dfrac{\overline{X}_{n_1} - \overline{Y}_{n_2} - \Delta}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$

1. Deducir un intervalo de confianza para $\Delta = \mu_1 - \mu_2$
2. Resolver el ejemplo. Vayamos al R.

Intervalo de confianza para la diferencia de medias

Sean $X_1, \ldots, X_{n_1}$ y $Y_1, \ldots, Y_{n_2}$ dos muestras aleatorias indendientes de las distribuciones $N\left(\mu_1, \sigma^2\right)$ y $N\left(\mu_2, \sigma^2\right)$ respectivamente. El intervalo

$$\left[\bar{X}_{n_1} - \bar{Y}_{n_2} - t_{n_1+n_2-2,\frac{\alpha}{2}} \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\,, \right.$$

$$\left. \bar{X}_{n_1} - \bar{Y}_{n_2} + t_{n_1+n_2-2,\frac{\alpha}{2}} \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right]$$

es un intervalo de confianza a nivel $1-\alpha$ para $\mu_1 - \mu_2$.