Test parametrici di confronto fra due popolazioni

Supponiamo di avere due popolazioni

$X_1, X_2, \dots, X_n$

$Y_1, Y_2, \dots, Y_m$ $\qquad n, m \in \mathbb{N}$

$$X \sim Nor\left(\mu = \mu_x,\ \sigma^2 = \sigma_x^2\right)$$

$$Y \sim Nor\left(\mu = \mu_y,\ \sigma^2 = \sigma_y^2\right)$$

Test

$$\text{Tipo I}: \begin{cases} H_0: \mu_X = \mu_Y \\ H_1: \mu_X \neq \mu_Y \end{cases}$$

$$\text{Tipo II}: \begin{cases} H_0: \mu_X = \mu_Y \quad \text{opp.} \quad \mu_X \leq \mu_Y \\ H_1: \mu_X > \mu_Y \end{cases}$$

$$\text{Tipo III}: \begin{cases} H_0: \mu_X = \mu_Y \quad \text{opp.} \quad \mu_X \geq \mu_Y \\ H_1: \mu_X < \mu_Y \end{cases}$$

Oss. che

$$\bar{X}_n \sim Nor\left(\mu = \mu_X,\ \sigma^2 = \frac{\sigma_x^2}{n}\right)$$

$$\bar{Y}_m \sim Nor\left(\mu = \mu_Y \;,\; \sigma^2 = \frac{\sigma_Y^2}{m}\right)$$

quindi $$\bar{X}_n - \bar{Y}_m \sim Nor\left(\mu = \mu_X - \mu_Y \;,\; \sigma^2 = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

Caso 1: Varianza nota

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim Nor(0, 1)$$

Allora $$U = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \overset{H_0}{\sim} Nor(0, 1)$$

Tipo	Rifiuto di H_0 se:
I	$\lvert u \rvert > z_{1-\frac{\alpha}{2}}$
II	$u > z_{1-\alpha}$
III	$u < -z_{1-\alpha}$

Esempio

Confronto delle prestazioni sportive prima e dopo le vacanze natalizie

prima delle vacanze 5 corse, tempo medio 56,41 sec.

dopo le vacanze 6 corse, tempo medio 58.82 sec.

Supponendo che il tempo X impiegato per percorrere 400 metri sia distribuito normalmente e che la varianza sia uguale ($\sigma^2 = 0.1$), le vacanze hanno contribuito a peggiorare le prestazioni

X = "tempo prima delle vacanze" $\sim Norm(\mu_X, \sigma^2 = 0.1)$

Y = "tempo dopo le vacanze" $\sim Nor(\mu_Y, \sigma^2 = 0.1)$

$H_0: \mu_X = \mu_Y$ le vacanze non hanno influito

$H_1: \mu_X < \mu_Y$ le vacanze hanno influito

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = -1.6611$$

$\alpha = 5\%$

$z_{1-\alpha} = 1.6449$

$u < -z_{1-\alpha}$ e quindi rifiuto H_0

Caso 2: Varianze incognite, uguali

$$\sigma_X^2 = \sigma_Y^2 = v$$

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2$$

$$S_Y^2 = \frac{1}{m-1} \sum_{i=1}^{m} \left(Y_i - \bar{Y}_m \right)^2$$

$$\hat{v} = S_p^2 = \frac{(n-1) S_X^2 + (m-1) S_Y^2}{n+m-2}$$

$\hat{v}$ prende il nome di varianza combinata

La statistica test è data da

$$U = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{S_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)}} \overset{H_0}{\sim} T(\nu = n+m-2)$$

Tipo	Rifiuto di H_0 se:
I	$\lvert u \rvert > t_{1-\frac{\alpha}{2}; n+m-2}$

II $\qquad u > t_{1-\alpha;\, n+m-2}$

III $\qquad u < -t_{1-\alpha;\, n+m-2}$

Caso 3: Varianze incognite, non uguali

$$V = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \overset{H_0}{\sim} T\left(\nu = \nu^*\right)$$

$$\nu^* = \text{P.I.}\left(\frac{\left(\frac{S_x}{n} + \frac{S_y}{m}\right)^2}{\frac{1}{n-1}\left(\frac{S_x^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{S_y^2}{m}\right)^2}\right)$$

parte intera

* Confronto di due Normali, a campioni accoppiati

Abbiamo un campione di ampiezza n, cioè formato da n coppie di osservazioni

$$(X_1, Y_1), (X_2, Y_2) \ldots, (X_n, Y_n)$$

X e Y sono normali e la loro differenza ($W = X - Y$) sarà anch'essa normale distribuita

$$W \sim Nor\left(\mu_W, \sigma_W^2\right)$$

con $\mu_W = \mu_X - \mu_Y$ e σ_W incognita

$$\text{tipo I}: \begin{cases} H_0: & \mu_W = 0 \\ H_1: & \mu_W \neq 0 \end{cases}$$

$$\text{tipo II}: \begin{cases} H_0: & \mu_W = 0 \quad \text{o} \quad \mu_W \leq 0 \\ H_1 & \mu_W > 0 \end{cases}$$

$$\text{tipo III}: \begin{cases} H_0: & \mu_W = 0 \quad \text{o} \quad \mu_W \geq 0 \\ H_1: & \mu_W < 0 \end{cases}$$

$$S_W^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(w_i - \overline{W}_n \right)$$

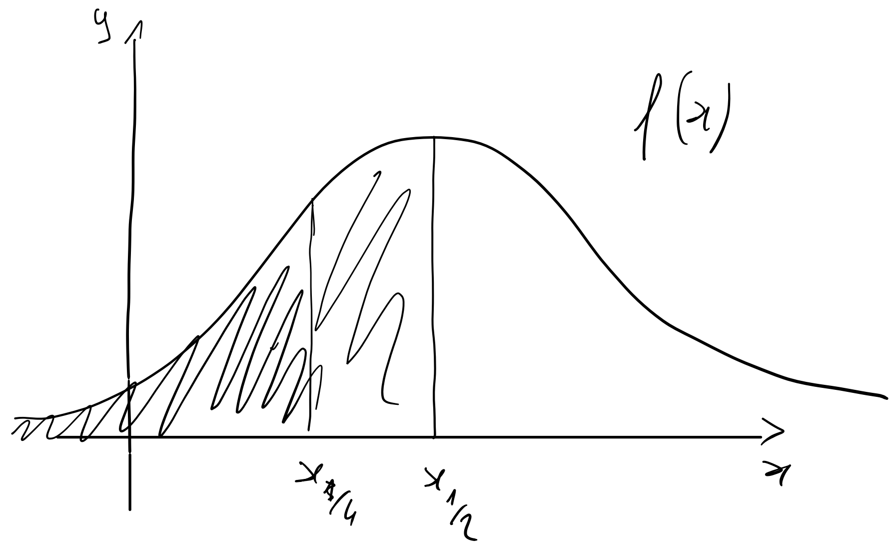
$$U = \frac{\overline{W}_n}{\sqrt{\frac{S_W^2}{n}}} \overset{H_0}{\sim} T\left(\nu = n-1\right)$$

Tipo	Rifiuto
I	$|u| > t_{1-\frac{\alpha}{2},\, n-1}$
II	$u > t_{1-\alpha,\, n-1}$
III	$u < -t_{1-\alpha,\, n-1}$

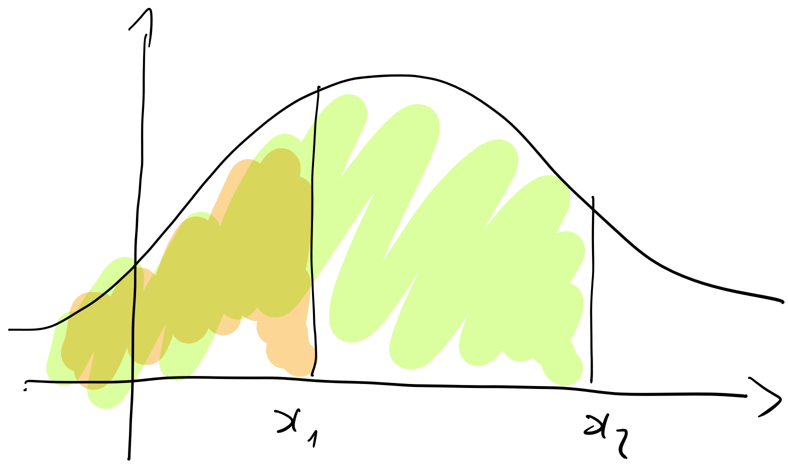
Esempio

paziente	1	2	3	4	5	6
tasso di colesterolo prima dell'esp.	170.3	162.2	210.3	157.3	180.1	192.3
tasso di colesterolo dopo l'esp.	166.5	165.8	171.1	163.2	152.2	161.5

Stabilire se il farmaco ha fatto effetto con livelli di significatività 1%, 5%, 10%



y
f(x)
x
$x_{1/4}$
$x_{1/2}$



x_1
x_2