Significatività della differenza tra percentuali nei g. c.

In alcuni problemi spesso si devono usare misure percentuali, invece di valori assoluti.

Se indichiamo con p'1 e p'2 le percentuali relative a due campioni basate su n1 ed n2 prove e se p1 e p2 rappresentano le percentuali delle corrispondenti popolazioni, allora la media e la deviazione standard delle differenze tra percentuali sono date dalle seguenti relazioni:

$$M_{p'_1 - p'_2} = p_1 - p_2$$

e

$$\sigma_{p'_1 - p'_2} = \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}}$$

in cui q1 = (1 - p'1) e q2 = (1 - p'2).

Il numero totale di mucche malate è 134, mentre complessivamente le mucche morte sono 11.

Quindi, avremo p=11/134=0,082 cioè 8,2% e q=1-0,082=0,918.

In questo problema p1 e p2 non sono noti.

Ipotesi H0
Se si suppone che p1 = p2 = p, sarà anche q1 = q2 = q e si avrà sotto H0:

$$M_{p'_1 - p'_2} = p_1 - p_2 = 0$$

<u>Esempio VII.8</u> Due diversi tipi di siero sono somministrati a 134 mucche colpite da una malattia.

- **Il siero A è iniettato a 50 mucche e 5 di queste muoiono.**
- **Il siero B è iniettato alle restanti 84 mucche e 6 di esse muoiono.**

Si vuol sapere se la differenza di effetto tra i due tipi di siero sulla mortalità delle mucche per la malattia in questione è significativa.

Il tasso di mortalità corrispondente al siero del tipo A è 5/50 = 1/10 ossia il 10% ovvero p'1 = 0,10.

Invece, il tasso di mortalità corrispondente al siero del tipo B è 6/84=1/14=0,0714 ossia il 7,14% ovvero p'2=0,0714.

$$\sigma_{p'_1 - p'_2} = \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}} = \sqrt{0{,}082x0{,}918x\left(\frac{1}{50} + \frac{1}{84}\right)}$$

cioè

$$\sigma_{p'_1 - p'_2} = \sqrt{\frac{0{,}082x0{,}918x134}{4200}} = \sqrt{0{,}002401663} = 0{,}049$$

Con questi risultati, si ottiene, quindi:

$$t = \frac{0{,}10 - 0{,}0714}{0{,}049} = \frac{0{,}0286}{0{,}049} = 0{,}58367$$

per cui:

$$\boxed{\frac{1}{\sqrt{2\pi}} \int_0^{0{,}58367} e^{-\frac{t^2}{2}} .dt = 0{,}2202.}$$

Quindi la probabilità che si abbia una differenza maggiore di (0,10-0,07)=0,03 è pari a (1-2.0,2202)=(1-0,4404)=0,5596.

Quindi la probabilità che si abbia una differenza maggiore di (0,10-0,07)=0,03 è pari a (1-2.0,2202)=(1-0,4404)=0,5596.
Si conclude, dunque, che circa 56 coppie di campioni casuali su 100 dovrebbero dare una differenza più grande di 0,03 cosicché la differenza tra i due sieri non è affatto significativa.

$\alpha = 0{,}1$

Regola di decisione :

Si rigetta l'ipotesi H0 se $t < -1{,}64$ o $t > 1{,}64$

$-1{,}64 < t = 0.58 < 1{,}64$ **si accetta l'ipotesi H0!**

Distribuzione della deviazione standard nei g.c.

$$\boxed{M_s = \sigma} \quad \text{e} \quad \boxed{\sigma_s = \frac{\sigma}{\sqrt{2n}}}$$

Inoltre, per grandi campioni la distribuzione della deviazione standard diviene approssimativamente normale.

Esempio VII.9

Si supponga di avere due grandi campioni per i quali si ha:

$n_1=329$; $n_2=302$;
$m_1=32{,}75$; $m_2=30{,}6$;
$s_1=8{,}05$; $s_2=6{,}95$.

Si supponga inoltre che sia $\sigma_1 = \sigma_2$.

Con queste posizioni applicando le formule indicate, si avrà:

$$z = \frac{(s_1 - s_2) - M_{s_1 - s_2}}{\sigma_{s_1 - s_2}} =$$

$$= \frac{(s_1 - s_2) - 0}{\sqrt{\sigma_{s_1}^2 + \sigma_{s_2}^2}} = \frac{(s_1 - s_2)}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$$

cioè

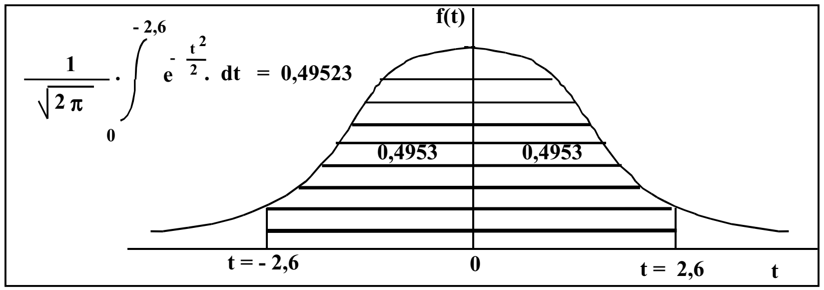
$$t = \frac{8{,}05 - 6{,}95}{\sqrt{\frac{8{,}05^2}{2x329} + \frac{6{,}95^2}{2x302}}} = \frac{1{,}10}{0{,}422439} = 2{,}6039.$$

Ora, dalla tavola della curva normale standardizzata si ha che per t = 2,6 l'integrale tra zero e t vale:

$$\boxed{\frac{1}{\sqrt{2\pi}}\int_0^{2,6} e^{-\frac{t^2}{2}}.dt = 0{,}4953}$$

e quindi:

$$1 - \frac{1}{\sqrt{2\pi}}\int_{-2,6}^{2,6} e^{-\frac{t^2}{2}}.dt = 1 - 2x0{,}4953 \approx 0{,}0094$$



Questa probabilità è più piccola del 5% ed infatti è 0,0094 < 0,05, per cui l'ipotesi che sia $\sigma 1 = \sigma 2$ non è giustificata. In altre parole, esiste una differenza significativa tra le deviazioni standard delle due popolazioni.

α = 0,05

Regola di decisione :

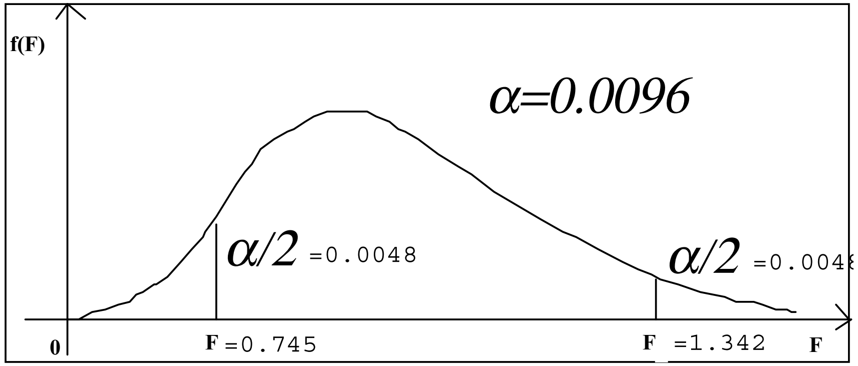
Si rigetta l'ipotesi H0 se t < - 1,96 o t>1,96

1.96< t=2,6 **si rifiuta l'ipotesi H0!**

Altro metodo di soluzione esatto!

$$\left.\begin{array}{l} H_0 : \sigma_x = \sigma_y \\ H_1 : \sigma_x \neq \sigma_y \end{array}\right\} \text{bilaterale} \qquad \frac{s_x^2}{\sigma_x^2}\frac{\sigma_y^2}{s_y^2} \approx F_{n_x-1, n_y-1}$$

$$\frac{s_x^2}{s_y^2} = \frac{8,05^2}{6.95^2} = 1,34 = F_{328,301} \qquad \frac{s_y^2}{s_x^2} = \frac{6.95^2}{8,05^2} = 0.745 = F_{301,328}$$



α = 0,05

Regola di decisione :

Si rigetta l'ipotesi H0 se

$$F < F_{\alpha=0.025, n_x-1, n_y-1} = \frac{1}{F_{1-\alpha=0.975, n_y-1, n_x-1}} = 0.801628$$

O:

$$F > F_{\alpha=0.975, n_x-1, n_y-1} = 1.248795$$

$$\frac{s_x^2}{s_y^2} = \frac{8,05^2}{6.95^2} = 1,34 = F_{328,301} > 1.2487$$

si rifiuta l'ipotesi H0!