http://geoappunti.altervista.org

Appunti
di
STATISTICA

Distribuzione esponenziale

X v.a. continua , $R_X = (0,+\infty)$

Si dice che X ha distribuzione esponenziale a parametro λ (numero reale) se la p.d.f. è così definita:

$$f_X = \begin{cases} \lambda\, e^{-\lambda x} & x>0 \\ 0 & \text{altrove} \end{cases}$$

$\lambda>0$
$E(X) = 1/\lambda$
$Var(X) = 1/\lambda^2$

Distribuzione Gamma

X v.a. continua , $R_X = (0,+\infty)$

Si dice che X ha distribuzione Gamma a parametro r, λ ε R se la p.d.f. è così definita:

$$f_X = \begin{cases} \dfrac{\lambda^2}{\Gamma(r)}\, x^{r-1}\, e^{-\lambda x} & x>0 \\ 0 & \text{altrove} \end{cases} \qquad r, \lambda>0$$

$E(X) = r/\lambda$
$Var(X) = r/\lambda^2$

Distribuzione chi-quadro

$r = \dfrac{n}{2} \quad n\geq 1$
$\lambda = ½$

X è una variabile aleatoria continua con supporto $R_X = (0,+\infty)$

Si dice che X ha distribuzione chi-quadro con n gradi di libertà se la densità di probabilità della X è:

$$f_X(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)}\, x^{n/2-1}\, e^{-x/2} & x>0 \\ 0 & \text{altrove} \end{cases}$$

$E(X) = n$
$Var(X) = 2n$

Distribuzione normale o Gaussiana

X è una variabile aleatoria continua con supporto $R_X = (-\infty,+\infty)$

Si dice che X ha distribuzione normale o Gaussiana a parametri $\mu \in R$ $\sigma^2>0$, se la p.d.f. è così costituita:

$$f_X(x) = \longrightarrow \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad x \in R$$

$X \sim N(\mu, \sigma^2)$

$E(X) = \mu$
$Var(X) = \sigma^2$
$\sigma = \sqrt{Var(X)}$ deviazione standard

Distribuzione t di Student

Sia:
$Z \sim N(0,1)$
$Y \sim \chi^2(n)$ $\quad n\geq 1$

Z, Y sono due v.a. indipendenti cioè i valori assunti da ciascuno di essi non influenzano l'altra.

$$T = \frac{Z}{\sqrt{Y/n}}$$

T è continua e il suo supporto coincide con tutto R e si dice che una variabile così definita ha distribuzione t di Student con n gradi di libertà e per brevità si scrive:

$$T \sim t(n) \qquad n\geq 1$$

Distribuzione di Fisher

$U \sim \chi^2(n_1)$

$n_1, n_2 \geq 1$

$V \sim \chi^2(n_2)$

U, V v.a. indipendenti ; $F = \dfrac{U}{n_1} : \dfrac{V}{n_2}$ è una v.a. continua con supporto $R_X = (0,+\infty)$.

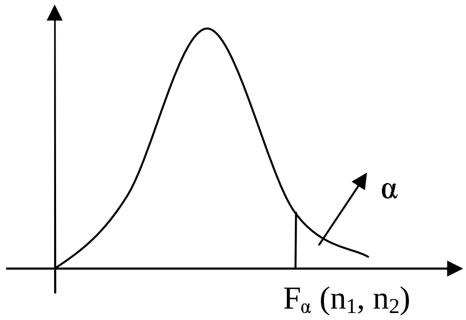
Si dice che F ha distribuzione di Fisher con n_1, n_2 gradi di libertà rispettivamente:

$F \sim F(n_1, n_2)$

$\dfrac{1}{F} \sim F(n_2, n_1)$

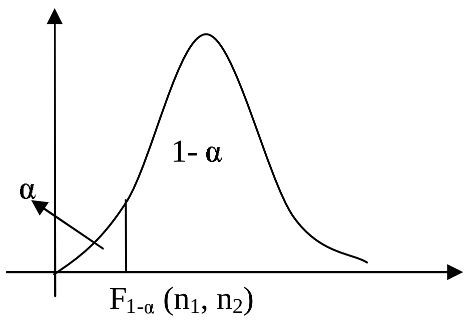
$p \equiv \alpha$

Percentile di ordine α di una $F \sim F(n_1, n_2)$, $F_\alpha(n_1, n_2)$



Se $p \equiv 1-\alpha$

Percentile di ordine $1-\alpha$ di $F \sim F(n_1, n_2)$, $F_{1-\alpha}(n_1, n_2)$



$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_1, n_2)}$$

http://geoappunti.altervista.org

Stima per intervallo

Supponiamo di avere solo una popolazione e di voler conoscere l'intervallo di fiducia:

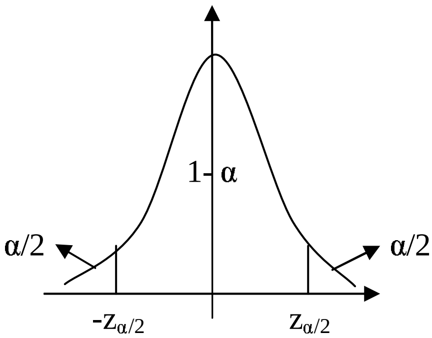
$X \sim B(1,p)$, $p \,\varepsilon\, (0,1)$

$H_0: p=p_0$ contro $H_1: p \neq p_0$

Fissato α si rigetta H_0 e si accetta H_1 se: $\left| Z = \dfrac{\overline{X} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \right| \geq z_{\alpha/2}$

$\alpha = P(\text{rigettare } H_0) = P(|Z_{calc}| \geq z_{\alpha/2})$

$1-\alpha = P(\text{accettare } H_0)$



$P(-z_{\alpha/2} \leq Z_{calc} \leq z_{\alpha/2}) \rightarrow P\left(-z_{\alpha/2} \leq \dfrac{\overline{X} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \leq z_{\alpha/2} \right)$

$\rightarrow P\left(\overline{X} - z_{\alpha/2}\sqrt{\dfrac{p_0(1-p_0)}{n}} \leq p_0 \leq \overline{X} + z_{\alpha/2}\sqrt{\dfrac{p_0(1-p_0)}{n}} \right)$

$\rightarrow$ Intervallo di fiducia a livello $(1-\alpha)$ per $p=\overline{X}$

$$\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}} , \overline{X} + z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}} \right)$$

$$\boxed{\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} , \overline{X} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} \right)}$$

Se $p_0 \,\varepsilon\, (\qquad)$ si accetta H_0

Se $p_0 \notin (\qquad)$ si rigetta H_0

Supponiamo di avere due popolazioni indipendenti e di voler conoscere l'intervallo di fiducia:

$X \sim B(1,p_1)$
$Y \sim B(1,p_2)$

$H_0 : p_1=p_2$ contro $H_1: p_1 \neq p_2$

Fissato α si rigetta H_0 e si accetta H_1 se $\left| \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{x(1-x)}{n} + \dfrac{y(1-y)}{m}}} \right| \geq z_{\alpha/2}$

$\rightarrow \alpha = P$ (rigettare H_0)
$\rightarrow 1-\alpha = P$ (accettare H_0)

$P(-z_{\alpha/2} \leq Z_{calc} \leq z_{\alpha/2}) \rightarrow P\left(-z_{\alpha/2} \leq \dfrac{\overline{X} - \overline{Y} - (p_1 - p_2)}{\sqrt{\dfrac{x(1-x)}{n} + \dfrac{y(1-y)}{m}}} \leq z_{\alpha/2}\right)$

$\rightarrow P\left(\overline{x} - \overline{y} - z_{\alpha/2}\sqrt{\dfrac{x(1-x)}{n} + \dfrac{y(1-y)}{m}} \leq p_1 - p_2 \leq \overline{x} - \overline{y} + z_{\alpha/2}\sqrt{\dfrac{x(1-x)}{n} + \dfrac{y(1-y)}{m}}\right)$

$\rightarrow$ Intervallo di fiducia a livello $(1-\alpha)$ per $p_1 - p_2$

$$\boxed{x - y - z_{\alpha/2}\sqrt{\frac{x(1-x)}{n} + \frac{y(1-y)}{m}}\ ,\ x - y + z_{\alpha/2}\sqrt{\frac{x(1-x)}{n} + \frac{y(1-y)}{m}}}$$

ESEMPI

1) Con numerose ricerche è stato dimostrato che un tossico diluito in acqua alla concentrazione standard determina la morte del 30% degli individui di una specie A.
Determinare un intervallo di fiducia al 95% per la proporzione di decessi su un campione di 80 individui.

$n=80 \quad p_0=30\%=0.3=\Sigma x_i/n$

p = proporzione di individui morti della specie A
X v.a. che rappresenta il risultato della singola prova
$X \sim B\,(1\text{-}p)$

$\alpha = 0.05 \rightarrow \alpha/2 = 0.025$
$Z = 0.025 = 1.96\%$

$$\left(0.3 - 1.96\sqrt{\frac{0.3\,(0.7)}{80}}\ ;\ 0.3 + 1.96\sqrt{\frac{0.3\,(0.7)}{80}}\right)$$

$$\left(0.3 - 0.1\ ;\ 0.3 + 0.1\right) = \left(0.2\ ;\ 0.4\right)$$

2) Un ricercatore deve verificare la differenza delle qualità delle falde idriche di due aree distinte. Analisi preliminari hanno dimostrato che nell'area 1 il 45% di n=130 prelievi supera i limiti di attenzione per almeno un parametro; mentre nell'area 2 tali limiti sono superati solo del 25% di m=130 prelievi. Testare a livello $\alpha = 0.05$

$H_0 = p_1 = p_2$ contro $H_1 = p_1 \neq p_2$
$X \sim B\,(1\text{-}p) \quad n=130$
$Y \sim B\,(1\text{-}p) \quad m=130$

$\Sigma\, x_i = 45\%\,(130) = 0.45\, n \rightarrow \bar{x} = \frac{1}{n}\,\Sigma\, x_i = 0.45 = \hat{p}_1$

$\Sigma\, y_j = 25\%\,(130) = 0.25\, m \rightarrow \bar{y} = \frac{1}{m}\,\Sigma\, y_j = 0.25 = \hat{p}_2$

$$\text{se } H_0 \text{ è vera : } Z = \frac{X - Y}{\sqrt{\frac{x(1\text{-}x)}{n} + \frac{y(1\text{-}y)}{m}}} \rightarrow \frac{0.45 - 0.25}{\sqrt{0.35\,(0.65)\,\frac{1}{260}}} = 3.33$$

Si rigetta H_0 se $Z_{calc} \geq z_{\alpha/2}$
$z_{\alpha/2} = 1.96$ quindi si rigetta H_0
Costruiamo un intervallo di fiducia per $p_1 - p_2$

$$\left(x - y - z_{\alpha/2}\ \frac{x(1\text{-}x)}{n} + \frac{y(1\text{-}y)}{m}\ ,\ x - y + z_{\alpha/2}\ \frac{x(1\text{-}x)}{n} + \frac{y(1\text{-}y)}{m}\right) = \left(0.08\ ;\ 0.32\right)$$

Intervallo di fiducia al 95%

Test per la media in modelli normali

X è una v.a. ~ $N(\mu, \sigma^2)$

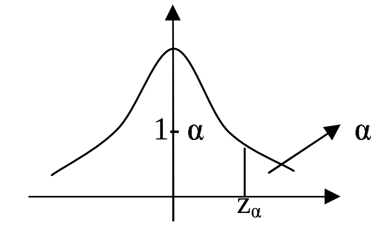
1°caso)
μ è incognita $\mu \in R$ $\mu_0 \in R$

H_0: $\mu = \mu_0$ contro H_1: $\mu > \mu_0$
$\mu < \mu_0$
$\mu \neq \mu_0$

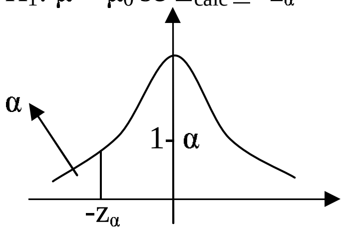
$$\overline{X} = \frac{1}{n} \Sigma X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Se H_0 è vera $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

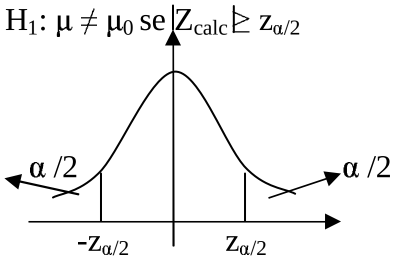
Fissato α si rigetta H_0 e si accetta
H_1: $\mu > \mu_0$ se $Z_{calc} \geq z_\alpha$



H_1: $\mu < \mu_0$ se $Z_{calc} \leq -z_\alpha$



H_1: $\mu \neq \mu_0$ se $|Z_{calc}| \geq z_{\alpha/2}$



$\rightarrow \alpha = P$ (rigettare H_0)
$\rightarrow 1- \alpha = P$ (accettare H_0)

$P\,(-z_{\alpha/2} < Z_{calc} < z_{\alpha/2}) \rightarrow P\left(-z_{\alpha/2} < \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right)$

$\rightarrow P\left(X - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} < \mu < X + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right)$

$\rightarrow$ Intervallo di fiducia a livello (1- α) per μ (σ^2 nota):

$$\boxed{\left[X - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ;\ X + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

Se $\mu_0 \in [\quad]$ si accetta H_0

Se $\mu_0 \notin [\quad]$ si rigetta H_0

2° caso)

σ^2 non è nota , n<30

$Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N\,(0,1)$

$S^2 = \dfrac{1}{n-1}\,\Sigma\,(X_i - \overline{X})^2$

Se $(X_1 \ldots X_n)$, $X_i \sim N(\mu, \sigma^2)$

$\rightarrow W = \dfrac{(n-1)\,S^2}{\sigma^2} \sim \chi^2\,(n-1)$

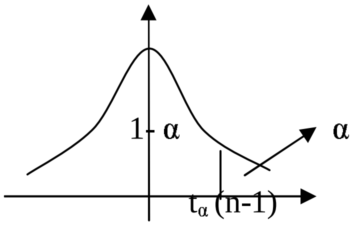
W, Z v.a. indipendenti

$T = \dfrac{Z}{\sqrt{W/n-1}}$ ~ t di Student con n-1 g.l.

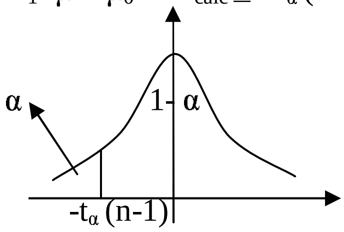
$$T = \frac{\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)\,S^2}{\sigma^2\,(n-1)}}} = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t\,(n-1)$$

Fissato α si rigetta H_0 e si accetta

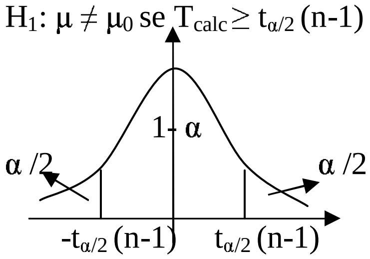
H_1: $\mu > \mu_0$ se $T_{calc} \geq t_\alpha (n-1)$



H_1: $\mu < \mu_0$ se $T_{calc} \leq -t_\alpha (n-1)$



H_1: $\mu \neq \mu_0$ se $T_{calc} \geq t_{\alpha/2} (n-1)$



→ $\alpha = P$ (rigettare H_0)

→ $1- \alpha = P$ (accettare H_0)

→ $P\,(-t_{\alpha/2} (n-1) < T < t_{\alpha/2} (n-1)$

→ $P \left(-t_{\alpha/2} (n-1) < \dfrac{\overline{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2} (n-1) \right)$

→ $P \left(\overline{X} - t_{\alpha/2} (n-1) \dfrac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2} (n-1) \dfrac{S}{\sqrt{n}} \right)$

→ Intervallo di fiducia a livello $(1- \alpha)$ per μ (σ^2 incognita $n<30$):

$$\boxed{\left[\overline{X} - t_{\alpha/2} (n-1) \frac{S}{\sqrt{n}} \,,\, \overline{X} + t_{\alpha/2} (n-1) \frac{S}{\sqrt{n}} \right]}$$

→ Intervallo di fiducia a livello $(1- \alpha)$ per σ^2:

Se $\mu_0 \in \left(\qquad \right)$ si accetta H_0

Se $\mu_0 \notin \left(\qquad \right)$ si rigetta H_0

3° caso)

σ^2 incognita , $n \geq 30$

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \approx N(0,1)$$

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$$

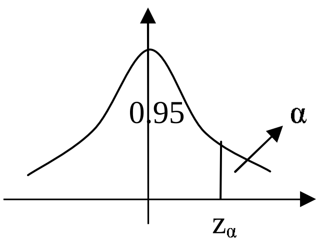
Possiamo notare come si ritorna al 1° caso. Se non è sufficientemente grande applichiamo il 3° caso se $X_i \not\sim N(\mu, \sigma^2)$ qualunque sia la distribuzione possiamo applicare un test per la media della distribuzione.

ESEMPI

1) $X \sim N(\mu\ , 0.36)$ di cui siano note le seguenti osservazioni:
0.8 – 1.8 – 1.0 – 0.1 – 0.9 – 1.7 – 1.0 – 1.1 – 0.9 – 1.2 – 0.5
Testare a livello $\alpha = 0.05$
$H_0 : \mu=1.5$ contro $H_1 : \mu<1.5$

Se H_0 è vera $Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \dfrac{x - 1.5}{0.6/\sqrt{11}} \sim N(0,1)$

Si rigetta H_0 se $Z_{calc} \leq -z_\alpha = -z_{0.05} = 1.65$
Se vogliamo calcolare la $P(Z\leq z)$ dove $z=0.04$ secondo la tabella il percentile è 0.5160



$0.95 = P(Z\leq z_{0.05}) = 1.65$
$\overline{x} = \dfrac{1}{11} \Sigma\, x_i = 1.027$

$Z = \dfrac{1.027 - 1.5}{0.6/\sqrt{11}} = -2.613 < -1.65 \rightarrow$ la statistica cade nella coda di sinistra per questo si rigetta H_0

2) Dall'esperienza passata si sa che il peso dei salmoni cresciuti in un allevamento commerciale ha distribuzione normale con media che varia da stagione a stagione e con deviazione standard sempre uguale a σ = 0.03 libbre. Quando grande occorre prendere il campione se vogliamo essere sicuri al 95 % che la nostra stima del peso medio dei salmoni di quest'anno sia precisa entro più o meno 0.1 libbra?

X : peso dei campioni in un anno
$X \sim N(\mu, (0.3)^2)$
$X_1 \ldots X_n$
Intervallo di fiducia per μ:
$$\left(X - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\ ;\ X + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

95% $\alpha = 0.05$ $1-\alpha = 0.95$
$\pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \approx \pm 0.1$ $Z_{0.025}=1.96$

dobbiamo cercare n tale che: $1.96 \dfrac{0.3}{\sqrt{n}} \leq 0.1$ $n\geq 34.57$

basta prendere un n>35 (salmoni) per poter essere sicuro al 95% che il peso medio non superi di 0.1 libbra dal valore centrale.

http://geoappunti.altervista.org

Test di verifica ipotesi per la varianza in modelli normali

$X \sim N(\mu,\sigma^2)$, $X_1 \ldots X_n$

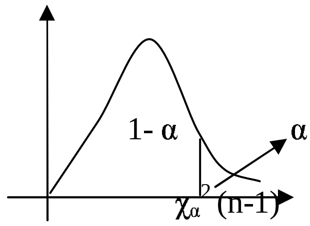
H_0: $\sigma^2 = \sigma^2_0$ contro H_1: $\sigma^2 > \sigma^2_0$
$\sigma^2 < \sigma^2_0$
$\sigma^2 \neq \sigma^2_0$

$\overline{X} = \frac{1}{n} \Sigma x_i$ stimatore di μ

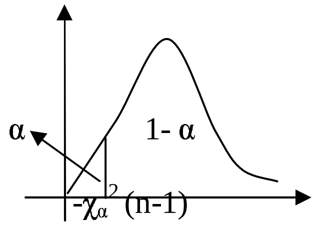
$S^2 = \frac{1}{n-1} \Sigma x_i$ stimatore di σ^2

$X_i \sim N(\mu,\sigma^2) \rightarrow W = \frac{(n-1)\, S^2}{\sigma^2} \sim \chi^2\,(n-1)$

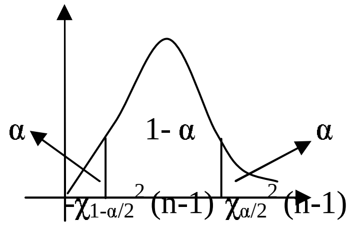
Fissato α si rigetta H_0 e si accetta
H_1: $\sigma^2 > \sigma^2_0$ se $W_{calc} \geq \chi_\alpha^2\,(n-1)$



Fissato α si rigetta H_0 e si accetta
H_1: $\sigma^2 < \sigma^2_0$ se $W_{calc} \leq \chi_\alpha^2\,(n-1)$



Fissato α si rigetta H_0 e si accetta
H_1: $\sigma^2 \neq \sigma^2_0$ se $W_{calc} \leq \chi_{1-\alpha/2}^2\,(n-1)$ oppure se $W_{calc} \geq \chi_{\alpha/2}^2\,(n-1)$



→ $\alpha = P$ (rigettare H_0)
→ $1-\alpha = P$ (accettare H_0)

→ $P\left(\chi^2_{1-\alpha/2}(n-1) < W < \chi^2_{\alpha/2}(n-1)\right)$

→ $P\left(\chi^2_{1-\alpha/2}(n-1) \leq \frac{(n-1)\,S^2}{\sigma^2} \leq \chi^2_{\alpha/2}(n-1)\right)$

→ $P\left(\frac{\chi^2_{1-\alpha/2}(n-1)}{(n-1)\,S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{\alpha/2}(n-1)}{(n-1)\,S^2}\right)$

→ $P\left(\frac{(n-1)\,S^2}{\chi^2_{\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$

→ Intervallo di fiducia a livello (1- α) per σ^2:

$$\boxed{\left(\frac{(n-1)\,S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)}$$

Se $\mu_0 \in \left(\quad\right)$ si accetta H_0

Se $\mu_0 \notin \left(\quad\right)$ si rigetta H_0

→ Intervallo di fiducia a livello (1- α) per σ:

$$\boxed{\left(\sqrt{\frac{(n-1)\,S^2}{\chi^2_{\alpha/2}(n-1)}}, \sqrt{\frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2}(n-1)}}\right)}$$

Test di verifica ipotesi per il confronto di varianze di due distribuzioni normali

Consideriamo due popolazioni indipendenti

$X \sim N(\mu,\sigma_1^2)$
$Y \sim N(\mu,\sigma_2^2)$

$H_0: \sigma_1^2 = \sigma_2^2$ contro $H_1: \sigma_1^2 > \sigma_2^2$
$\sigma_1^2 < \sigma_2^2$
$\sigma_1^2 \neq \sigma_2^2$

$\rightarrow H_0: \dfrac{\sigma_1^2}{\sigma_2^2} = 1$ contro $H_1: \dfrac{\sigma_1^2}{\sigma_2^2} > 1$
$\dfrac{\sigma_1^2}{\sigma_2^2} < 1$
$\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$

$\overline{X}_1 = \dfrac{1}{n} \Sigma x_i$ stimatore di μ_1

$S_X^2 = \dfrac{1}{n-1} \Sigma (x_i - \overline{x})^2$ stimatore di σ_1^2

$\rightarrow W = \dfrac{(n-1)\, S_X^2}{\sigma_1^2} \sim \chi^2\,(n-1)$

$\rightarrow$ v.a. indipendenti

$\overline{Y}_1 = \dfrac{1}{n} \Sigma x_j$ stimatore di μ_2

$S_Y^2 = \dfrac{1}{n-1} \Sigma (x_j - \overline{x})^2$ stimatore di σ_2^2

$\rightarrow V = \dfrac{(m-1)\, S_Y^2}{\sigma_2^2} \sim \chi^2\,(m-1)$

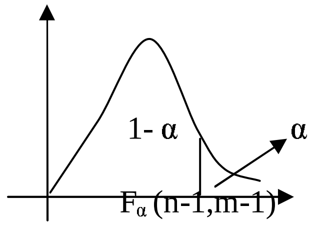
$$F = \frac{\dfrac{W}{n-1}}{\dfrac{V}{m-1}} \sim F(n-1,\, m-1)$$

$$F = \frac{\dfrac{\cancel{(n-1)}\, S_X^2}{\sigma_1^2\, \cancel{(n-1)}}}{\dfrac{\cancel{(m-1)}\, S_Y^2}{\sigma_2^2\, \cancel{(m-1)}}} = \frac{S_X^2}{S_Y^2}\, \frac{\sigma_2^2}{\sigma_1^2}$$

se H_0 è vera : $F = \dfrac{S_X^2}{S_Y^2}$

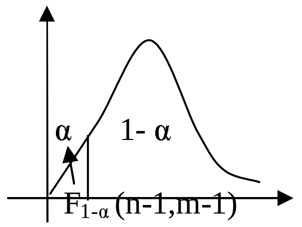
Fissato α si rigetta H_0 e si accetta

H_1: $\sigma_1^2 > \sigma_2^2$ se $F_{calc} \geq F_\alpha (n-1,m-1)$



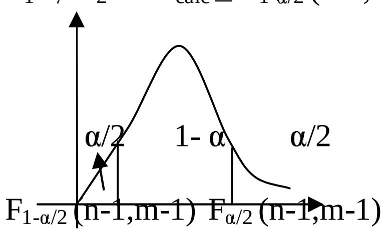
Fissato α si rigetta H_0 e si accetta

H_1: $\sigma_1^2 < \sigma_2^2$ se $F_{calc} \leq F_{1-\alpha} (n-1,m-1)$



Fissato α si rigetta H_0 e si accetta

H_1: $\sigma_1^2 \neq \sigma_2^2$ se $F_{calc} \leq F_{1-\alpha/2} (n-1,m-1)$ oppure se $F_{calc} \geq F_{\alpha/2} (n-1,m-1)$



$\rightarrow \alpha = P$ (rigettare H_0)

$\rightarrow 1-\alpha = P$ (accettare H_0)

$$\rightarrow P \left[F_{1-\alpha/2} (n-1,m-1) \leq F \leq F_{\alpha/2} (n-1,m-1) \right]$$

$$\rightarrow P \left[F_{1-\alpha/2} (n-1,m-1) \leq \frac{S_X^2}{S_Y^2} \frac{\sigma_2^2}{\sigma_1^2} \leq F_{\alpha/2} (n-1,m-1) \right]$$

$$\rightarrow P \left[\frac{S_Y^2}{S_X^2} F_{1-\alpha/2} (n-1,m-1) \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{S_Y^2}{S_X^2} F_{\alpha/2} (n-1,m-1) \right]$$

$\rightarrow$ Intervallo di fiducia a livello $(1-\alpha)$ per $\frac{\sigma_2^2}{\sigma_1^2}$:

$$\boxed{\frac{S_Y^2}{S_X^2} F_{1-\alpha/2} (n-1,m-1) , \frac{S_Y^2}{S_X^2} F_{\alpha/2} (n-1,m-1)} \rightarrow F_{1-\alpha/2} (n-1,m-1) = \frac{1}{F_{\alpha/2} (n-1,m-1)}$$

Test parametrici di intervalli di fiducia per la differenza di media in modelli normali

1) Campioni indipendenti:

Abbiamo due popolazioni distinte:
Per la prima popolazione avremo: $X \sim N(\mu_x, \sigma^2_x)$
Per la seconda popolazione avremo: $Y \sim N(\mu_y, \sigma^2_y)$
Entrambe le popolazioni sono indipendenti

$H_0: \mu_x = \mu_y$ contro $H_1: \mu_x > \mu_y$
$\mu_x < \mu_y$
$\mu_x \neq \mu_y$

a) σ^2_x e σ^2_y sono note

$\overline{X} = \frac{1}{n} \Sigma X_i$ stimatore di μ_x

$\overline{Y} = \frac{1}{m} \Sigma Y_j$ stimatore di μ_y

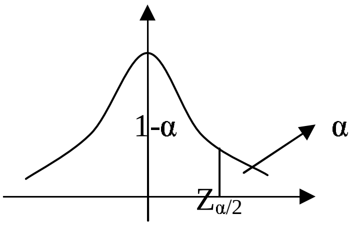
$\overline{X} \sim N(\mu_x, \frac{\sigma^2_x}{n})$
$\rightarrow$ sono indipendenti
$\overline{Y} \sim N(\mu_y, \frac{\sigma^2_y}{m})$

$$\rightarrow \overline{X} - \overline{Y} \sim N\left(\mu_x - \mu_y ; \frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}\right)$$

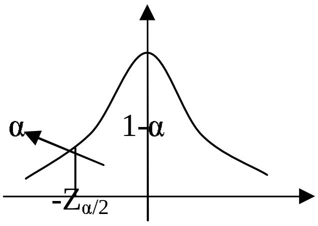
$$\rightarrow Z = \frac{\overline{x} - \overline{y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}}} \rightarrow Z \sim N(0,1)$$

se H_0 è vera : $Z = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}}}$

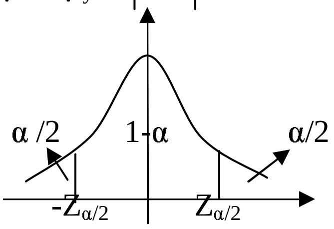
Fissato α si rigetta H_0 e si accetta
H_1: $\mu_x > \mu_y$ se $Z_{calc} \geq Z_{\alpha/2}$



Fissato α si rigetta H_0 e si accetta
H_1: $\mu_x < \mu_y$ se $Z_{calc} \leq -Z_{\alpha/2}$



Fissato α si rigetta H_0 e si accetta
H_1: $\mu_x \neq \mu_y$ se $|Z_{calc}| \geq Z_{\alpha/2}$



$\rightarrow \alpha = P$ (rigettare H_0)
$\rightarrow 1-\alpha = P_\theta$ (accettare H_0) $\qquad \theta = \mu_x - \mu_y$

$\rightarrow P_\theta \left(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\right)$

$$\rightarrow P_\theta \left(-Z_{\alpha/2} \leq \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}}} \leq Z_{\alpha/2} \right)$$

$$\rightarrow P_\theta \left(\bar{x} - \bar{y} - Z_{\alpha/2}\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}} \leq (\mu_x - \mu_y) \leq \bar{x} - \bar{y} + Z_{\alpha/2}\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}} \right)$$

$\rightarrow$ Intervallo di fiducia a livello $(1-\alpha)$ per $\mu_x - \mu_y$:

$$\boxed{\bar{x} - \bar{y} - Z_{\alpha/2}\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}} \; ; \; \bar{x} - \bar{y} + Z_{\alpha/2}\sqrt{\frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}}}$$

b) σ^2_x e σ^2_y sono incognite quindi n,m<30

$\sigma^2_x = \sigma^2_y$

$$\left.\begin{array}{l} \overline{X} \sim N\left(\mu_x, \dfrac{\sigma^2_x}{n}\right) \\ \overline{Y} \sim N\left(\mu_y, \dfrac{\sigma^2_y}{m}\right) \end{array}\right\} \rightarrow \text{sono v.a. indipendenti}$$

$$\left.\begin{array}{ll} S_X^2 = \dfrac{1}{n-1} \Sigma (x_i - \overline{x})^2 & \text{stimatore di } \sigma^2_x \\ S_Y^2 = \dfrac{1}{m-1} \Sigma (y_j - \overline{y})^2 & \text{stimatore di } \sigma^2_y \end{array}\right\}$$

$$\left.\begin{array}{l} W = \dfrac{(n-1)\, S_X^2}{\sigma^2_x} \sim \chi^2\,(n-1) \\ \\ V = \dfrac{(m-1)\, S_Y^2}{\sigma^2_y} \sim \chi^2\,(m-1) \end{array}\right\} \rightarrow \text{v.a. indipendenti}$$

$$\rightarrow X - Y \sim N\left(\mu_x - \mu_y\,;\ \frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m}\right)$$

$$\rightarrow Z = \frac{\overline{x} - \overline{y} - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma^2_x}{n} + \dfrac{\sigma^2_y}{m}}} \rightarrow Z \sim N\,(0,1)$$

$$U = \frac{(n-1)\, S_X^2}{\sigma^2_x} + \frac{(m-1)\, S_Y^2}{\sigma^2_y} \sim \chi^2\,(n+m-2)$$

$$T = \frac{\dfrac{x - y - (\mu_x - \mu_y)}{\sqrt{\sigma^2 \left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}}{\sqrt{\dfrac{(n-1)\, S_X^2 + (m-1)\, S_Y^2}{(n+m-2)}}} = \frac{x - y - (\mu_x - \mu_y)}{\sqrt{S^2_p \left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$$

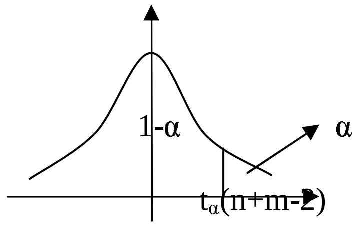
$$S^2_p = \frac{(n-1)\, S_X^2 + (m-1)\, S_Y^2}{n+m-2} \qquad \text{varianza campionaria associata}$$

$$T = \frac{x - y - (\mu_x - \mu_y)}{S_p \sqrt{\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$$

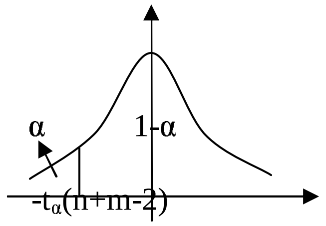
$T \sim t\,(n+m–2)$

se H_0 è vera : $T = \dfrac{x - y - (\mu_x - \mu_y)}{S_p \sqrt{\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$

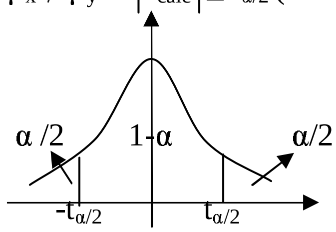
Fissato α si rigetta H_0 e si accetta
H_1: $\mu_x > \mu_y$ se $T_{calc} \geq t_\alpha(n+m-2)$



Fissato α si rigetta H_0 e si accetta
H_1: $\mu_x < \mu_y$ se $T_{calc} \leq t_\alpha(n+m-2)$



Fissato α si rigetta H_0 e si accetta
H_1: $\mu_x \neq \mu_y$ se $|T_{calc}| \geq t_{\alpha/2}(n+m-2)$



$\rightarrow \alpha = P$ (rigettare H_0)
$\rightarrow 1-\alpha = P_\theta$ (accettare H_0) $\qquad \theta = \mu_x - \mu_y$

$\rightarrow P_\theta \left(-t_{\alpha/2}(n+m-2) \leq T \leq t_{\alpha/2}(n+m-2)\right)$

$$\rightarrow P_\theta \left(-t_{\alpha/2}(n+m-2) \leq \frac{x - y - (\mu_x - \mu_y)}{S_p \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \leq t_{\alpha/2}(n+m-2) \right)$$

$$\rightarrow P_\theta \left((x - y) - t_{\alpha/2}(n+m-2)\, S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \leq \mu_x - \mu_y \leq (x - y) + t_{\alpha/2}(n+m-2)\, S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

→ Intervallo di fiducia a livello (1- α) per $\mu_x - \mu_y$:

$$\left((\bar{x} - \bar{y}) - t_{\alpha/2}(n+m-2)\, S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \;;\; (\bar{x} - \bar{y}) + t_{\alpha/2}(n+m-2)\, S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

c) σ^2_x e σ^2_y sono incognite quindi n,m<30

$\sigma^2_x \neq \sigma^2_y$

$$T' = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}}} \approx t([\upsilon])$$

$$\upsilon = \frac{\left(\frac{S^2_x}{n} + \frac{S^2_y}{m}\right)^2}{\frac{\left(\frac{S^2_x}{n}\right)^2}{(n-1)} + \frac{\left(\frac{S^2_y}{m}\right)^2}{(m-1)}}$$

Se H_0 e vera : $T' = \dfrac{\bar{x} - \bar{y}}{\frac{S^2_x}{n} + \frac{S^2_y}{m}}$

Fissato α si rigetta H_0 e si accetta H_1 se: $\mu_x > \mu_y$ se $T_{calc} \geq t_\alpha(\upsilon)$

$\mu_x < \mu_y$ se $T_{calc} \leq t_\alpha(\upsilon)$

$\mu_x \neq \mu_y$ se $|T_{calc}| \geq t_{\alpha/2}(\upsilon)$

→ $\alpha = P$ (rigettare H_0)

→ $1- \alpha = P_\theta$ (accettare H_0) $\theta = \mu_x - \mu_y$

→ $P_\theta\left(-t_{\alpha/2}(\upsilon) \leq T' \leq t_{\alpha/2}(\upsilon)\right)$

→ $P_\theta \; -t_{\alpha/2}(\upsilon) \leq \dfrac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}}} \leq t_{\alpha/2}(\upsilon)$

→ $P_\theta\left((\bar{x} - \bar{y}) - t_{\alpha/2}(\upsilon)\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}} \leq \mu_x - \mu_y \leq (\bar{x} - \bar{y}) + t_{\alpha/2}(\upsilon)\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}} \right)$

→ Intervallo di fiducia a livello (1- α) per $\mu_x - \mu_y$:

$$\boxed{(x - y) - t_{\alpha/2}(\upsilon)\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}} \quad ; \quad (x - y) + t_{\alpha/2}(\upsilon)\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}}}$$

d) σ^2_x e σ^2_y sono incognite $n,m \geq 30$

$X \not\sim N(\mu_x,\sigma^2_x)$

$Y \not\sim N(\mu_y,\sigma^2_y)$

$$Z = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}}}$$

Per il T.L.C.

$Z \approx N(0,1)$

Questo ci riporta al caso “a”

→ Intervallo di fiducia a livello (1- α) per $\mu_x - \mu_y$:

$$\boxed{(x - y) - z_{\alpha/2}(\upsilon)\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}} \quad ; \quad (x - y) + z_{\alpha/2}(\upsilon)\sqrt{\frac{S^2_x}{n} + \frac{S^2_y}{m}}}$$