$\bar{X} = \frac{\sum_1^n x_i}{n}$ $\quad S^2 = \frac{\sum_1 (x - \bar{x})^2}{n-1}$ $\quad Var(x) = E(x^2) - E^2(x)$

momento secondo

momento primo al ^2

media al quadrato

$MSE = Var(\hat{\theta}) = ?$

So che per $P(X \leq k) = F(x)$ approx con Th limite centrale → media $\Phi\left(\frac{k - \mu}{\sigma / \sqrt{n}}\right)$

Formulario per l'esame di Statistica Allievi INF, TEL. AA 08/09 Docente: Ilenia Epifani

Probabilità

Var(αX) = α2 Var(X)
Var(X + β) = Var(X)
Var(X) = E(X2) − E(X)2

Variabili aleatorie discrete

Nome simbolo parametri	Densità discreta $p_X(k) = P[X = k]$	F.g.m. $m_X(t) = E(e^{tX})$	Media	Varianza
bernoulliana $X \sim \mathbf{Be}(p) = \mathbf{Bi}(1,p)$ $0 < p < 1$	$p^k(1-p)^{1-k}$ $k = 0, 1$ p = prob. di $x_i = 1$	$(1-p) + pe^t$	p	$p(1-p)$
binomiale $X \sim \mathbf{Bi}(n,p)$ $0 < p < 1,\ n = 1, 2, \ldots$	$\binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, \ldots, n$	$(1 - p + pe^t)^n$	np	$np(1-p)$
ipergeometrica $(b+r, r, n)$ $b, r \geq 0, n \geq 1$ interi $n \leq b + r$	$\frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}$ $\max\{0, n-b\} \leq k \leq \min\{r, n\}$ k intero		$\frac{nr}{b+r}$	$\frac{nrb}{(b+r)^2}\left(1 - \frac{n-1}{b+r-1}\right)$
geometrica (p) $0 < p < 1$ $f(x,p) = p(1-p)^{x-1}$	$p_X(k) = p(1-p)^{k-1}$ $k = 1, 2, \ldots$ $\boxed{k = x}$	$\frac{pe^t}{1 - (1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
di Poisson $X \sim \mathcal{P}(\lambda)$ $\lambda > 0$ $\boxed{k = x}$	$\frac{e^{-\lambda}\lambda^k}{k!}$ $k = 0, 1, \ldots$	$\exp[\lambda(e^t - 1)]$	λ	λ
uniforme su $1, 2, \ldots, n$ $n \geq 1$ intero	$\frac{1}{n}$ $k = 1, 2, \ldots, n$	$\frac{e^t(1 - e^{nt})}{n(1 - e^t)}$	$\frac{n+1}{2}$	$\frac{n^2 - 1}{12}$

Sono uguali!

$P(X_0 \leq k) = \sum_0^k f(x) = \sum_0^k \frac{e^{-\lambda}\lambda^k}{k!}$

Variabili aleatorie continue

Nome simbolo parametri	Densità continua $f_X(x) = F'_X(x)$ $\forall x$ t.c. F'_X esiste	F.g.m.	Media	Varianza
uniforme $X \sim \mathcal{U}(a,b)$ $a < b$	$\frac{1}{b-a}\mathbf{1}_{(a,b)}(x)$	$\begin{cases} \frac{e^{bt} - e^{at}}{(b-a)t} & \text{se } t \neq 0 \\ 1 & \text{se } t = 0 \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
normale [gaussiana] $X \sim N(\mu, \sigma^2)$ $\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	$\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$	μ	σ^2
lognormale $\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$ $(x > 0)$		$e^{\mu + \sigma^2/2}$	$e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)$
gamma $X \sim \Gamma(a, \beta)$ $a, \beta > 0$	$\frac{1/\beta^a}{\Gamma(a)} x^{a-1} e^{-x/\beta} \mathbf{1}_{(0,+\infty)}(x)$	$\frac{1}{(1-\beta t)^a}$ $\forall t < 1/\beta$	$a\beta$	$a\beta^2$
esponenziale $X \sim \mathcal{E}(\beta)$ $\beta > 0$	$\mathcal{E}(\beta) = \Gamma(1, \beta)$	$\frac{1}{1 - \beta t}$ $\forall t < 1/\beta$	β	β^2
chiquadro $X \sim \chi^2_n$ $n \geq 1$, intero	$\chi^2_n = \Gamma(n/2, 2)$	$\left(\frac{1}{1-2t}\right)^{n/2}$ $\forall t < \frac{1}{2}$	n	$2n$
Weibull $\alpha, \beta > 0$	$\frac{\alpha}{\beta} x^{\alpha - 1} e^{-x^\alpha/\beta} \mathbf{1}_{(0,\infty)}(x)$		$\Gamma(1 + 1/\alpha)\beta^{1/\alpha}$	$\beta^{2/\alpha}[\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)]$
t di Student $X \sim t_n$ $n = 1, \ldots$	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$		0 se $n > 1$	$\frac{n}{n-2}$ se $n > 2$

$f(y|\theta) = \frac{1}{\theta} e^{-y/\theta}$

Integrale gamma: $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$ e $\Gamma(a+1) = a\Gamma(a)$ $\forall a > 0$

Vettore gaussiano bivariato $(X, Y)^T \sim \mathcal{N}(\mu_X, \mu_Y, \sigma^2_X, \sigma^2_Y, \varrho)$ $\left[\varrho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}\right]$

$$f_{X\,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1 - \varrho^2}} e^{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

$Var(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2$

$Cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)\,E(Y)$

f.d.r bidimensionale continua: $F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t)\, ds\, dt$

f.d.r bidimensionale discreta: $F_{X,Y}(x, y) = \sum_{s \leq x} \sum_{t \leq y} p_{X,Y}(s, t)$

Densità marginali di (X, Y) continuo: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$

Densità marginali di (X, Y) discreto: $p_X(x) = \sum_{y_k} p_{X,Y}(x, y_k)$

$X \sim \mathcal{E}(\lambda)$ $\quad E(x) = \frac{1}{\lambda}$

$\int \beta x\, e^{\alpha x} dx = \frac{\beta e^{\alpha x}(\alpha x - 1)}{\alpha^2}$

Esponenziale: $f(x) = \lambda e^{-\lambda x}$

$E(x) = \frac{1}{\lambda}$ $\quad Var(x) = \lambda^{-2}$

$Var(\bar{x}) = \frac{\sigma^2}{n}$ $\|$ → var media campionaria

$F_X(x) = P(X \leq x)$

$\bar{x} = \frac{\sum x_i}{n} \approx \mu$

questi
$s_0^2 = \frac{1}{n} \sum (x_i - \mu)^2$ SE CONOSCO LA MEDIA
se ho la media vera

Formulario per l'esame di Statistica Allievi INF TEL. AA 07/08 Docente: Ilenia Epifani
Statistica

1 Test di ipotesi sulla media di una popolazione gaussiana

$(x_1, \ldots, x_n)$ = realizzazione campionaria di $X_1, \ldots, X_n$ *i.i.d.* $\sim N(\mu, \sigma^2)$.

σ^2 **nota [z-test]:** → TEST × GRANDI CAMPIONI

H₀	H₁	Si rifiuta H₀ se	p-value
$\mu=\mu_0$ $\mu=\mu_0$ $\mu\le\mu_0$	$\mu=\mu_1$ con $\mu_0<\mu_1$ $\mu>\mu_0$ $\mu>\mu_0$	$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \ge z_{1-\alpha}$	$1-\Phi\left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right)$
$\mu=\mu_0$ $\mu=\mu_0$ $\mu\ge\mu_0$	$\mu=\mu_1$ con $\mu_0>\mu_1$ $\mu<\mu_0$ $\mu<\mu_0$	$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \le -z_{1-\alpha}$	$\Phi\left(\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right)$
$\mu=\mu_0$	$\mu\ne\mu_0$	$\frac{\lvert\bar{x}-\mu_0\rvert}{\sigma/\sqrt{n}} \ge z_{1-\frac{\alpha}{2}}$	$2\left[1-\Phi\left(\frac{\lvert\bar{x}-\mu_0\rvert}{\sigma/\sqrt{n}}\right)\right]$

→ x I.C. ±

Se hai diff. di 2 var $\sim N$, $\sigma = \sqrt{VAR(\bar{x}) + VAR(\bar{y})} = \sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}$; così e denominatore non va $\sqrt{n}$

Stimatore MLE anche è $\mu_{MLE} = \bar{x}$

Stimatore MLE della varianza $\sigma^2_{MLE} = \frac{(n-1)s^2}{n}$

STIMATORI MLE GAUSS

σ^2 **incognita [t-test]:** Stimatore varianza campionaria s: $s_n^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$

H₀	H₁	Si rifiuta H₀ se	p-value
$\mu=\mu_0$ $\mu=\mu_0$ $\mu\le\mu_0$	$\mu=\mu_1$ con $\mu_0<\mu_1$ $\mu>\mu_0$ $\mu>\mu_0$	$\frac{\bar{x}-\mu_0}{s/\sqrt{n}} \ge t_{n-1}(1-\alpha)$	$1-P\left(T_{n-1} \le \frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right)$
$\mu=\mu_0$ $\mu=\mu_0$ $\mu\ge\mu_0$	$\mu=\mu_1$ con $\mu_0>\mu_1$ $\mu<\mu_0$ $\mu<\mu_0$	$\frac{\bar{x}-\mu_0}{s/\sqrt{n}} \le -t_{n-1}(1-\alpha)$	$P\left(T_{n-1} \le \frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right)^{*} = \bar{\alpha}$ → minimo α per cui accetto H0
$\mu=\mu_0$	$\mu\ne\mu_0$	$\frac{\lvert\bar{x}-\mu_0\rvert}{s/\sqrt{n}} \ge t_{n-1}(1-\frac{\alpha}{2})$	$2\left[1-P\left(T_{n-1} \le \frac{\lvert\bar{x}-\mu_0\rvert}{s/\sqrt{n}}\right)\right]$

$* \; P(T_{11} \le -3) = F_{11}(-3) = 1 - F_{11}(3)$

$\bar{x}$ = media campionaria di $x_1, \ldots, x_n$
s^2 = varianza campionaria di $x_1, \ldots, x_n$
Φ = f.d.r. $N(0,1)$ e z_p t.c. $\Phi(z_p) = p$
$T_{n-1} \sim t$ di student con $n-1$ gradi di libertà e $t_{n-1}(p)$ t.c. $P(T_{n-1} \le t_{n-1}(p)) = p$.

T	m	p	$\hat{F}_0$	Fo	$\hat{F}$ -	L(x)	U(x)
	1	1/8	1/2				
	1	1/6	3/6				

TEST DI BUON ADATTAMENTO (KOLMOGOROV-SMIRNOFF)

Funz. di ripartiz. empirica $\hat{F}_n$.
Dato un campione 1, 0, 1, -1, 3, 2.5, 3, 1 $n=8$

$$\hat{F}_8(x) = \begin{cases} 0 & x < -1 \\ 1/8 & -1 \le x < 0 \\ 2/8 & 0 \le x < 1 \\ 5/8 & 1 \le x < 2.5 \\ 6/8 & 2.5 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

Prendo tutti i valori < 2,5

TEST calcolo $\overset{A}{|\hat{F}_n(x_i) - F_0(x_i)|}$ e ~~$\overset{B}{|\hat{F}_n(x_{i-1}) - F_0(x_i)|}$~~ e faccio tabellina. Trovo per ~~$D_n = \sup(A \cup B)$~~
$D_n = \sup(A)$ perciò D_n tra tutti gli n

Rifiuto H0 se D_n è "GRANDE", cioè se $P_{H_0}(D_n > K) = \alpha \rightsquigarrow K = q_{\alpha,n}$ // tabella tabella *
Nella tabella, dato α e n, trovo K.
se $D_n < K$ ACCETTO H0
se $D_n > K$ RIFIUTO H0 a significatività α

$\boxed{\begin{matrix} H_0: F = F_0 \\ H_1: F \ne F_0 \end{matrix}}$

* Se non hai α, pensa $\alpha = 10\%$

BANDE DI CONFIDENZA al $\alpha\%$ ($\alpha = 90\%$) Uso i quantili con $1-\alpha$ $90\% \to 10\%$
$[L(x)\ U(x)]$ $L(x) = \max(\hat{F} - q_{\alpha,n}, 0)$
$U(x) = \min(\hat{F} + q_{\alpha,n}, 1)$

Neyman - Pearson

$\frac{L(\theta_0)}{L(\theta_1)} \le \delta \rightsquigarrow x \le k \rightsquigarrow P(x \le k) = \alpha \rightsquigarrow \int_0^k f(x) = \alpha$

$f(x, \theta_0)$

TEST DI IPOTESI

2 χ^2-test sulla varianza di una popolazione gaussiana

μ **nota:**

momento secondo → momento primo al ^{2}

Varianza campionaria: $E(x^2) - E^2(x)$

H_0	H_1	Si rifiuta H_0 se	p-value
$\sigma^2=\sigma_0^2$ $\sigma^2=\sigma_0^2$ $\sigma^2\le\sigma_0^2$	$\sigma^2=\sigma_1^2$ *con* $\sigma_0^2<\sigma_1^2$ $\sigma^2>\sigma_0^2$ $\sigma^2>\sigma_0^2$	$\frac{ns_0^2}{\sigma_0^2} \ge \chi_n^2(1-\alpha)$	$1 - F_n\left(\frac{ns_0^2}{\sigma_0^2}\right)$
$\sigma^2=\sigma_0^2$ $\sigma^2=\sigma_0^2$ $\sigma^2\ge\sigma_0^2$	$\sigma^2=\sigma_1^2$ *con* $\sigma_0^2>\sigma_1^2$ $\sigma^2<\sigma_0^2$ $\sigma^2<\sigma_0^2$	$\frac{ns_0^2}{\sigma_0^2} \le \chi_n^2(\alpha)$	$F_n\left(\frac{ns_0^2}{\sigma_0^2}\right)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{ns_0^2}{\sigma_0^2} \ge \chi_n^2(1-\frac{\alpha}{2})$ oppure $\frac{ns_0^2}{\sigma_0^2} \le \chi_n^2(\frac{\alpha}{2})$	$2\min\{p_1,p_2\}$ dove $p_1=F_n\left(\frac{ns_0^2}{\sigma_0^2}\right)$ e $p_2=1-p_1$

$s_0^2 = (1/n)\sum_{j=1}^n (x_j-\mu)^2$.

μ **incognita:**

H_0	H_1	Si rifiuta H_0 se	p-value
$\sigma^2=\sigma_0^2$ $\sigma^2=\sigma_0^2$ $\sigma^2\le\sigma_0^2$	$\sigma^2=\sigma_1^2$ *con* $\sigma_0^2<\sigma_1^2$ $\sigma^2>\sigma_0^2$ $\sigma^2>\sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2} \ge \chi_{n-1}^2(1-\alpha)$	$1 - F_{n-1}\left(\frac{(n-1)s^2}{\sigma_0^2}\right)$
$\sigma^2=\sigma_0^2$ $\sigma^2=\sigma_0^2$ $\sigma^2\ge\sigma_0^2$	$\sigma^2=\sigma_1^2$ *con* $\sigma_0^2>\sigma_1^2$ $\sigma^2<\sigma_0^2$ $\sigma^2<\sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2} \le \chi_{n-1}^2(\alpha)$	$F_{n-1}\left(\frac{(n-1)s^2}{\sigma_0^2}\right)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{(n-1)s^2}{\sigma_0^2} \ge \chi_{n-1}^2(1-\frac{\alpha}{2})$ o $\frac{(n-1)s^2}{\sigma_0^2} \le \chi_{n-1}^2(\frac{\alpha}{2})$	$2\min\{p_1,p_2\}$ dove $p_1=F_{n-1}\left(\frac{(n-1)s^2}{\sigma_0^2}\right)$ e $p_2=1-p_1$

USA QUESTA → $Q = \sum_i \sum_j \left(\underbrace{n_{ij\,oss}}_{1^a \text{ tabella}} - \underbrace{n_{ij\,H_0}}_{2^a \text{ tabella}} \right)^2 / n_{ij\,H_0}$

ZONE IN GRIGIO NELLE TABELLE (n_{ij})

F_n = funzione di ripartizione chi-quadro con n gradi di libertà e $\chi_n^2(p)$ t.c. $F_n(\chi_n^2(p)) = p$.

TEST χ^2 DI INDIPENDENZA → NO se ho GAUSSIANE BIVARIATE!

statistica test: $Q = n \sum_{i=1}^{\#x} \sum_{j=1}^{\#y} \frac{N_{ij}^2}{N_i \cdot N_j} - n = q_{oss}$

DATI OSSERVATI:

$J = \#Y$

$X \backslash Y$				N_i
COND 1				•
COND 2				•
N_J	•	•	•	

$i = \#X$

Σ righe

Σ colonne

campioni ($\Sigma N_i = \Sigma N_j$)

Rifiuto H_0 se $Q > K$. $P(Q > k) = \alpha$

② H_0

$X \backslash Y$	C			N_j
COND 1				A
COND 2				
N_J	B			T

$= \frac{AB}{T} = C$

tabella generata se sono indipendenti

H_0: X e Y sono indipendenti

H_1: X e Y non sono indipendenti

$Q \sim \chi^2_{(i-1)(j-1)}$ lo individuo con $F_{(i-1)(j-1)}$

in uscita ho area %.

p-value $= 1 - F_{(i-1)(j-1)}(q_{oss})$

se $\begin{cases} \text{p-val} > 10\% & \text{SONO IND.} \\ \text{p-val} < 10\% & \text{RIFIUTO } H_0 \end{cases}$

$\boxed{\text{USA } 1-\alpha}$

TEST χ^2 DI ADATTAMENTO ($X \sim F$?)

H_0: $X \sim F$

H_1: $X \nsim F$

campioni $n \ge 50$ x' asintotico

	COND 1	COND 2	COND 3
N_i			
p_{0i}	"95"		
$n\,p_{0i}$			

$p_{0i} = P(\text{COND } i)$ dato H_0

{ quanti dovrebbero essere i casi su n e p_i

STAT TEST: $Q_n = \sum_{i=1}^{\#x} \frac{(N_i - np_{0i})^2}{np_{0i}} = \sum \frac{N_i^2}{np_{0i}}$

RIFIUTO H_0 se

$Q_n > \chi^2_{\#x-1-s}(1-\alpha)$ → % rifiuto H_0

CONDIZIONI - 1

s = # parametri da stimare

$s = 0$ se non ho parametri

se ti dà di usare questi

p-value. prendi Q_n, poi su tavola, trova valori α e poi fai $1-\alpha$

3 Test per il confronto di medie di due popolazioni gaussiane

3.1 Caso di campioni indipendenti:

$(x_1, \ldots, x_m)$ = realizzazione di $\boldsymbol{X} = X_1, \ldots, X_m$ *i.i.d.* $\sim N(\mu_X, \sigma_X^2)$,
$(y_1, \ldots, y_n)$ = realizzazione di $\boldsymbol{Y} = Y_1, \ldots, Y_n$ *i.i.d.* $\sim N(\mu_Y, \sigma_Y^2)$ e $\boldsymbol{X}, \boldsymbol{Y}$ indipendenti.

σ_X^2, σ_Y^2 note [z-test]:

$\mathbf{H_0}$	$\mathbf{H_1}$	**Si rifiuta $\mathbf{H_0}$ se**	**p-value**
$\mu_X = \mu_Y + \Delta$ $\mu_X \le \mu_Y + \Delta$	$\mu_X > \mu_Y + \Delta$ $\mu_X > \mu_Y + \Delta$	$\dfrac{\bar{x} - \bar{y} - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \ge z_{1-\alpha}$	$1 - \Phi\left(\dfrac{\bar{x} - \bar{y} - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}\right)$
$\mu_X = \mu_Y + \Delta$ $\mu_X \ge \mu_Y + \Delta$	$\mu_X < \mu_Y + \Delta$ $\mu_X < \mu_Y + \Delta$	$\dfrac{\bar{x} - \bar{y} - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \le -z_{1-\alpha}$	$\Phi\left(\dfrac{\bar{x} - \bar{y} - \Delta}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}\right)$
$\mu_X = \mu_Y + \Delta$	$\mu_X \neq \mu_Y + \Delta$	$\dfrac{\lvert\bar{x} - \bar{y} - \Delta\rvert}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \ge z_{1-\frac{\alpha}{2}}$	$2\left[1 - \Phi\left(\dfrac{\lvert\bar{x} - \bar{y} - \Delta\rvert}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}\right)\right]$

$\sigma_X^2 = \sigma_Y^2$ incognite ma uguali [t-test]:

$\mathbf{H_0}$	$\mathbf{H_1}$	**Si rifiuta $\mathbf{H_0}$ se**	**p-value**
$\mu_X = \mu_Y + \Delta$ $\mu_X \le \mu_Y + \Delta$	$\mu_X > \mu_Y + \Delta$ $\mu_X > \mu_Y + \Delta$	$\dfrac{\bar{x} - \bar{y} - \Delta}{s_p\sqrt{1/m + 1/n}} \ge t_{m+n-2}(1-\alpha)$	$1 - P\left(T_{m+n-2} \le \dfrac{\bar{x} - \bar{y} - \Delta}{s_p\sqrt{1/m + 1/n}}\right)$
$\mu_X = \mu_Y + \Delta$ $\mu_X \ge \mu_Y + \Delta$	$\mu_X < \mu_Y + \Delta$ $\mu_X < \mu_Y + \Delta$	$\dfrac{\bar{x} - \bar{y} - \Delta}{s_p\sqrt{1/m + 1/n}} \le -t_{m+n-2}(1-\alpha)$	$P\left(T_{m+n-2} \le \dfrac{\bar{x} - \bar{y} - \Delta}{s_p\sqrt{1/m + 1/n}}\right)$
$\mu_X = \mu_Y + \Delta$	$\mu_X \neq \mu_Y + \Delta$	$\dfrac{\lvert\bar{x} - \bar{y} - \Delta\rvert}{s_p\sqrt{1/m + 1/n}} \ge t_{m+n-2}(1-\frac{\alpha}{2})$	$2 - 2P\left(T_{m+n-2} \le \dfrac{\lvert\bar{x} - \bar{y} - \Delta\rvert}{s_p\sqrt{1/m + 1/n}}\right)$

$s_p^2 = \frac{s_X^2(m-1) + s_Y^2(n-1)}{m+n-2}$ con s_X^2 = varianza campionaria di $x_1, \ldots, x_m$ e s_Y^2 = varianza campionaria di $y_1, \ldots, y_n$,
$T_{m+n-2} \sim t$ di student con $m + n - 2$ gradi di libertà.

3.2 Caso di campioni accoppiati:

$(X_1, Y_1), \ldots, (X_n, Y_n)$ *i.i.d.* $\sim N\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \varrho\sigma_X\sigma_Y \\ \varrho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right)$ e $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \varrho$ incogniti. Per i problemi di verifica di ipotesi:

- $H_0 : \mu_X \le \mu_Y + \Delta$ contro $H_1 : \mu_X > \mu_Y + \Delta$
- $H_0 : \mu_X \ge \mu_Y + \Delta$ contro $H_1 : \mu_X < \mu_Y + \Delta$
- $H_0 : \mu_X = \mu_Y + \Delta$ contro $H_1 : \mu_X \neq \mu_Y + \Delta$

svolgere opportuno t-test usando il campione $X_1 - Y_1, \ldots, X_n - Y_n$.

$$\alpha = P_{\{\mu_D = 0\}} \frac{\sqrt{n} \cdot \bar{D}}{\sqrt{S^2_D}}$$

→ verifica distribuzione

TEST DI OMOGENEITÀ DI WILCOXON - MANN - WHITNEY

H_0: $F(x) = G(x)$ campione x(0.80, 1.83, 1.45, 1.31, 0.73) $m=5$ (ranghi 3, 7, 6, 8, 1)

H_1: $F(x) \gtreqless G(x)$ $G \sim y$(1.15, 0.90, 0.74) $n=3$ (ranghi 5, 4, 2)

$T_x = \sum$ ranghi di x = $3 + 7 + 6 + 8 + 1 = 25$

SE H_1: $F(x) \le G(x)$ $w_{1-\alpha} = m(m+n+1) - w_{\alpha}$ tabulato $1-p$

Rifiuto H_0 e accetto H_1 se $T_x > w_{1-\alpha}$

SE H_1: $F(x) \ge G(x)$

$w_\alpha = m(m+n+1) - w_{1-p}$ → Uso formula solo se α non è tabulato

Rifiuto H_0 e accetto H_1 se $T_x < w_\alpha$

SE: H_1: $F(x) \neq G(x)$

rifiuto H_0 e accetto H_1 se

$T_x < w_{\alpha/2}$ o $T_x > w_{1-\alpha/2}$

4 F-test per il confronto di varianze di due popolazioni gaussiane

$(x_1, \dots, x_m)$ = realizzazione di $\boldsymbol{X} = X_1, \dots, X_m$ *i.i.d.* $\sim N(\mu_X, \sigma_X^2)$,
$(y_1, \dots, y_n)$ = realizzazione di $\boldsymbol{Y} = Y_1, \dots, Y_n$ *i.i.d.* $\sim N(\mu_Y, \sigma_Y^2)$ e $\boldsymbol{X}, \boldsymbol{Y}$ indipendenti.

μ_X, μ_Y **note:**

Non è la Chi-quadr
è la Fisch (b)

$\mathbf{H_0}$	$\mathbf{H_1}$	**Si rifiuta $\mathbf{H_0}$ se**	**p-value**
$\sigma_X^2 \le \sigma_Y^2$	$\sigma_X^2 > \sigma_Y^2$	$\frac{s_{0X}^2}{s_{0Y}^2} \ge F_{m,n}(1-\alpha)$	$1 - P\left(F_{m,n} \le \frac{s_{0X}^2}{s_{0Y}^2}\right)$
$\sigma_X^2 \ge \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$	$\frac{s_{0X}^2}{s_{0Y}^2} \le F_{m,n}(\alpha)$	$P\left(F_{m,n} \le \frac{s_{0X}^2}{s_{0Y}^2}\right)$
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 \ne \sigma_Y^2$	$s_{0X}^2/s_{0Y}^2 \ge F_{m,n}(1-\alpha/2)$ oppure $s_{0X}^2/s_{0Y}^2 \le F_{m,n}(\alpha/2)$	$2\min\{p_1,p_2\}$ dove $p_1 = P(F_{m,n} \le s_{0X}^2/s_{0Y}^2)$ e $p_2 = 1 - p_1$

$s_{0X}^2 := \frac{\sum_{j=1}^m (x_j - \mu_X)^2}{m} = \frac{\sum_i x_i^2 - m\mu_X^2}{m}$ e $s_{0Y}^2 := \frac{\sum_{j=1}^n (y_j - \mu_Y)^2}{n}$.

Se uso $\bar{x}$ e non μ (medie non note) $m' = m-1$ $n' = n-1$ (come prima Gauss)

$F_{a,b}$ = v.a. avente densità di Fisher con (a,b) gradi di libertà e $F_{a,b}(p)$ t.c. $P(F_{a,b} \le F_{a,b}(p)) = p$.

$$\mathrm{Var}(\bar{x}) = \frac{\bar{x}^2}{m} - \left(\frac{\bar{x}}{m}\right)^2 \rightarrow \text{meno peso al }{}^2$$

↳ Ramo sec.

(*)

μ_X, μ_Y **incognite:**

$\mathbf{H_0}$	$\mathbf{H_1}$	**Si rifiuta $\mathbf{H_0}$ se**	**p-value**
$\sigma_X^2 \le \sigma_Y^2$	$\sigma_X^2 > \sigma_Y^2$	$\frac{s_X^2}{s_Y^2} \ge F_{m-1,n-1}(1-\alpha)$	$1 - P\left(F_{m-1,n-1} \le \frac{s_X^2}{s_Y^2}\right)$
$\sigma_X^2 \ge \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$	$\frac{s_X^2}{s_Y^2} \le F_{m-1,n-1}(\alpha)$	$P\left(F_{m-1,n-1} \le \frac{s_X^2}{s_Y^2}\right)$
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 \ne \sigma_Y^2$	$s_X^2/s_Y^2 \ge F_{m-1,n-1}(1-\alpha/2)$ oppure $s_X^2/s_Y^2 \le F_{m-1,n-1}(\alpha/2)$	$2\min\{p_1,p_2\}$ dove $p_1 = P(F_{m-1,n-1} \le s_X^2/s_Y^2)$ e $p_2 = 1 - p_1$

TEST DI INDIPENDENZA PER BIVARIATE GAUSS

Date X, Y e $(X,Y) \sim N_{\text{bivariata}}$ → TEST INDIPENDENZA

1) impostiamo coeff. di correl. lineare di (X,Y) ρ

$$\begin{cases} H_0: \rho \le 0 \\ H_1: \rho > 0 \end{cases}$$

2) stimiamo ρ con:

$$R = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{j=1}^{n}(y_i - \bar{y})^2}} = \frac{\mathrm{Cov}_{X,Y}}{\sqrt{S_X^2 S_Y^2}}$$

covarianza campionaria

varianza campionaria

Rifiuto H_0 se

$$t := \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} \ge t_{n-2}(1-\alpha)$$