STATISTICA

VERIFICA D’IPOTESI - 3

Due campioni **indipendenti**

Confrontiamo la spesa media annua per riparazioni dell'auto tra uomini e donne

Confrontiamo la longevità tra una popolazione isolana e una continentale

Confrontiamo l'effetto di un farmaco tra il gruppo del trattamento e quello del placebo

Confrontiamo l'insorgere di una certa malattia tra chi fa sport e chi fa vita sedentaria

Due campioni **indipendenti**

Confrontiamo la spesa media annua per riparazioni dell’auto tra uomini e donne

Confrontiamo la longevità tra una popolazione isolana e una continentale

Confrontiamo l’effetto di un farmaco tra il gruppo del trattamento e quello del placebo

Confrontiamo l’insorgere di una certa malattia tra chi fa sport e chi fa vita sedentaria

due gruppi di dati da due popolazioni **indipendenti**

Esempio 1

	Infarto miocardico			$\widehat{p}_i$
Gruppo	**Sì**	**No**		
1-Placebo	189	10845	11034	0.0171
2-Aspirina	104	10933	11037	0.0094

$$H_0 : p_1 = p_2 \quad , \quad \boldsymbol{H_1 : p_1 > p_2}$$

Verifica d'ipotesi: p_1 e p_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $bern(p_1)$ $(n_1 p_1 \geq 5\ ecc.)$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $bern(p_2)$ $(n_2 p_2 \geq 5\ ecc.)$

$\boldsymbol{H_0}$	$\boldsymbol{H_1}$	Rifiutiamo $\boldsymbol{H_0}$ se:
$p_1 = p_2$	$p_1 \neq p_2$	$\left\lvert \dfrac{\hat{\boldsymbol{p}}_1 - \hat{\boldsymbol{p}}_2}{\sqrt{una\ varianza\ opportuna}} \right\rvert > z_{1-\alpha/2}$
$p_1 = p_2$	$p_1 > p_2$	$\dfrac{\hat{\boldsymbol{p}}_1 - \hat{\boldsymbol{p}}_2}{\sqrt{una\ varianza\ opportuna}} > z_{1-\alpha}$
...	...	...

Verifica d'ipotesi: p_1 e p_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $bern(p_1)$ $(n_1 p_1 \geq 5 \; ecc.)$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $bern(p_2)$ $(n_2 p_2 \geq 5 \; ecc.)$

Sotto $\boldsymbol{H_0}$:

$$\bar{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

stima della proporzione totale di "successi"

$$una \; varianza \; opportuna \; : \; \bar{p}\,(1-\bar{p}\,)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Verifica d’ipotesi: p_1 e p_2

L’IPOTESI DI INDIPENDENZA

DEI DUE CAMPIONI

GIOCA NEL CALCOLO DELLA VARIANZA *OPPORTUNA*

$$una\ varianza\ opportuna\ :\ \bar{p}\,(1-\bar{p}\,)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

Verifica d'ipotesi: p_1 e p_2

$(X_1, \ldots, X_{n_1})$ campione aleatorio $bern(p_1)$ $\quad (n_1 p_1 \geq 5\ ecc.)$

$(Y_1, \ldots, Y_{n_2})$ campione aleatorio $bern(p_2)$ $\quad (n_2 p_2 \geq 5\ ecc.)$

$\boldsymbol{H_0}$	$\boldsymbol{H_1}$	Rifiutiamo $\boldsymbol{H_0}$ se:
$p_1 = p_2$	$p_1 \neq p_2$	$\left\lvert \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\,(1-\bar{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \right\rvert > z_{1-\alpha/2}$
$p_1 = p_2$	$p_1 > p_2$	$\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\,(1-\bar{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} > z_{1-\alpha}$
…	…	…

Esempio 1

	Infarto miocardico			$\hat{p}_i$
Gruppo	**Sì**	**No**		
1-Placebo	189	10845	11034	0.0171
2-Aspirina	104	10933	11037	0.0094

$$H_0 : p_1 = p_2 \quad , \quad \boldsymbol{H_1 : p_1 > p_2}$$

Facciamo un salto in R

Script4a.R

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\overline{p}\,(1-\overline{p}\,)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con dev. standard di 3.2 kg.

Sottoporre a verifica l’ipotesi nulla che lo sport non abbia effetto sul peso.

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con dev. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

$\left(X_1, \dots, X_{n_1}\right)$ i.i.d $\sim N(\mu_1, {\sigma_1}^2)$ **camp. indip.** $(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_2, {\sigma_2}^2)$

"sportivi" "sedentari"

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con dev. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

$\left(X_1, \dots, X_{n_1}\right)$ i.i.d $\sim N(\mu_1, {\sigma_1}^2)$ **camp. indip.** $(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_2, {\sigma_2}^2)$

"sportivi" "sedentari"

$\boldsymbol{H_0 : \mu_1 = \mu_2}$ $\boldsymbol{H_1 : \mu_1 \neq \mu_2}$

Verifica d'ipotesi: μ_1 e μ_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $N(\mu_1, {\sigma_1}^2)$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $N(\mu_2, {\sigma_2}^2)$

$\boldsymbol{{\sigma_1}^2 = {\sigma_2}^2}$ **ma non note**

$\boldsymbol{H_0}$	$\boldsymbol{H_1}$	Rifiutiamo $\boldsymbol{H_0}$ se:
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left\lvert \dfrac{\boldsymbol{\bar{x}_1 - \bar{x}_2}}{\sqrt{una\ varianza\ opportuna}} \right\rvert > t(n_1 + n_2 - 2)_{1-\alpha/2}$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\boldsymbol{\bar{x}_1 - \bar{x}_2}}{\sqrt{una\ varianza\ opportuna}} > t(n_1 + n_2 - 2)_{1-\alpha}$
…	…	…

Verifica d'ipotesi: μ_1 e μ_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $N(\mu_1, {\sigma_1}^2)$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $N(\mu_2, {\sigma_2}^2)$

$\boldsymbol{{\sigma_1}^2 = {\sigma_2}^2}$ **ma non note**

$$\text{"}una\ varianza\ opportuna\text{"} \ : \ {s_p}^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

(*pooled*)

con
$$ {s_p}^2 = \frac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{n_1 + n_2 - 2}$$

Verifica d'ipotesi: μ_1 e μ_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $N(\mu_1, \sigma^2)$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $N(\mu_2, \sigma^2)$

$$s_p{}^2 = \frac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}$$

H_0	H_1	Rifiutiamo H_0 se:
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left\lvert \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p{}^2 \times \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \right\rvert > t(n_1 + n_2 - 2)_{1-\alpha/2}$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p{}^2 \times \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} > t(n_1 + n_2 - 2)_{1-\alpha}$
…	…	…

Verifica d'ipotesi: μ_1 e μ_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $E(X_i) = \mu_1, Var(X_i) = \sigma^2$ $\quad n_1 > 30$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $E(Y_i) = \mu_2, Var(Y_i) = \sigma^2$ $\quad n_2 > 30$

H_0	H_1	Rifiutiamo H_0 se:
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left\lvert \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{{s_p}^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right\rvert > t(n_1 + n_2 - 2)_{1-\alpha/2}$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{{s_p}^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > t(n_1 + n_2 - 2)_{1-\alpha}$
…	…	…

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

"sportivi"

$(X_1, \dots, X_{n_1})$ i.i.d $\sim N(\mu_1, \sigma^2)$

camp. indip.

"sedentari"

$(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_1, \sigma^2)$

FATE un salto in R

e scrivete la funzione t.medie

Esempio 2: valori soluzione

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

"sportivi"

$\left(X_1, \ldots, X_{n_1}\right)$ i.i.d $\sim N(\mu_1, \sigma^2)$

camp. indip.

"sedentari"

$\left(Y_1, \ldots, Y_{n_2}\right)$ i.i.d $\sim N(\mu_1, \sigma^2)$

$$\left|\frac{\bar{x}_1 - \bar{x}_2}{s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| = \frac{|71.2 - 70|}{2.81 \times \sqrt{\frac{1}{50} + \frac{1}{35}}} = 1.94$$

$$\boldsymbol{H_0 : \mu_1 = \mu_2} \qquad \boldsymbol{H_1 : \mu_1 \neq \mu_2}$$

cfr. $t(83)_{0.975} \approx 1.9901$

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

"sportivi"

$(X_1, \dots, X_{n_1})$ i.i.d $\sim N(\mu_1, \boldsymbol{\sigma_1}^2)$

camp. indip.

"sedentari"

$(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_1, \boldsymbol{\sigma_2}^2)$

Se non è possibile assumerle uguali, c'è una soluzione, che agisce sui gradi di libertà della distribuzione t, usualmente implementata nei software statistici.

Ese

Il peso me
svolger
pales .
standa

Il peso **ta**
sende 70 kg con
devi. sta

Sottoporre a ver port non abbia effetto sul peso.

Si potrebbe fare un test

$H_0 : \sigma_1^2 = \sigma_2^2$

contro

$H_0 : \sigma_1^2 \neq \sigma_2^2$**?**

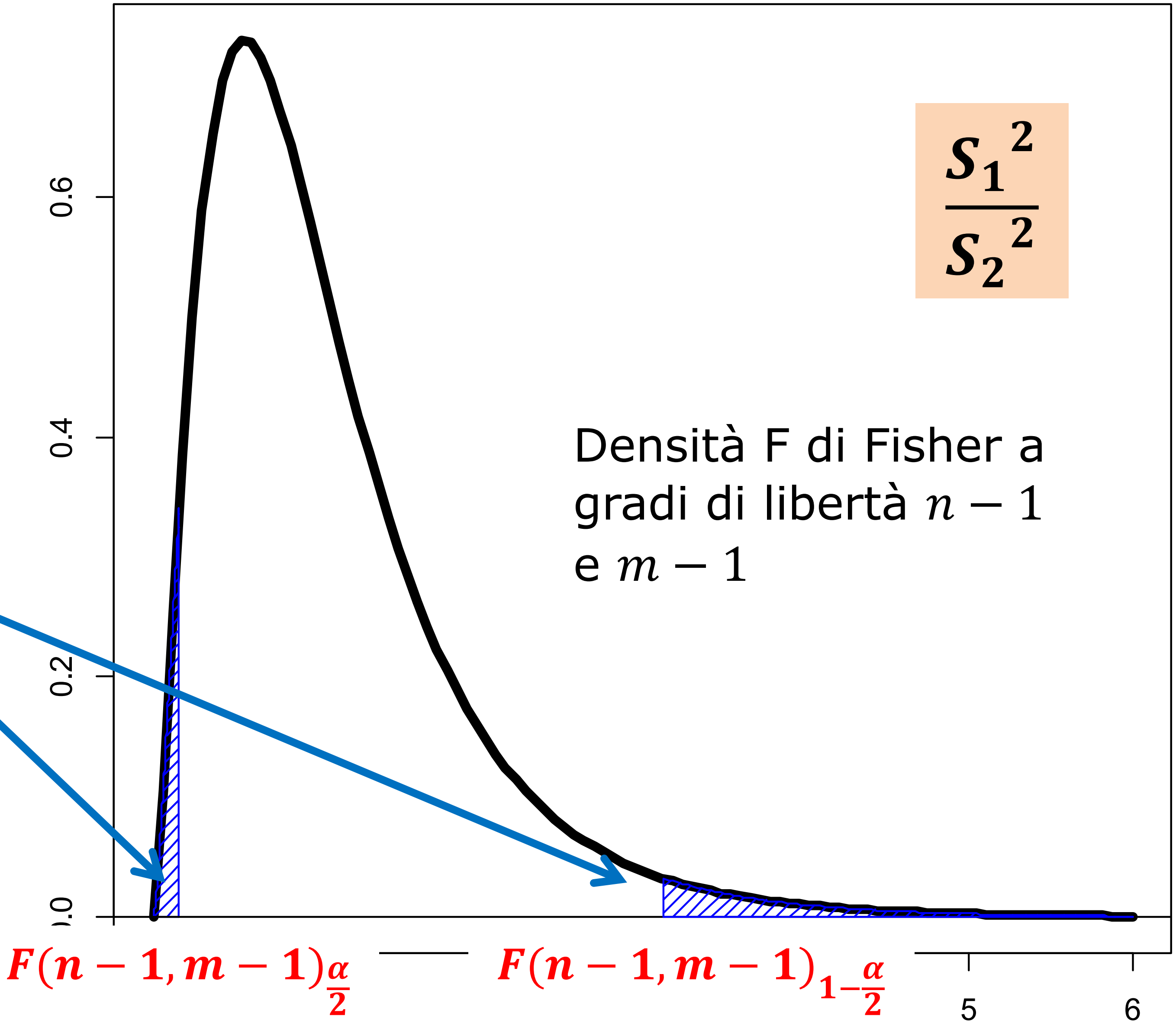
"sportivi"		"sedentari"
$(X_1, \dots, X_{n_1})$ i.i.d $\sim N(\mu_1, \sigma_1^2)$	**camp. indip.**	$(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_1, \sigma_2^2)$

Se non è possibile assumerle uguali, c'è una soluzione, che agisce sui gradi di libertà della distribuzione t, usualmente implementata nei software statistici.

Test sulle varianze

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$



Un ***contro***esempio

Effetto di una sostanza omeopatica contro l'ipertensione:

Un campione casuale di individui di cui misuriamo la pressione **prima** e **dopo** la somministrazione del farmaco.

Individuo	Pres. min. PRIMA	Pres. min. DOPO
A. Bianchini	83	75
B. Rossini	90	87
C. Verdini	66	62
E. Giallini	85	86
F. Turchini	72	70
…	…	…

Un ***contro***esempio

Effetto di una sostanza omeopatica contro l'ipertensione:

Un campione casuale di individui di cui misuriamo la pressione **prima** e **dopo** la somministrazione del farmaco.

Individuo	Pres. min. PRIMA	Pres. min. DOPO
A. Bianchini	83	75
B. Rossini	90	87
C. Verdini	66	62
E. Giallini	85	86
F. Turchini	72	70
…	…	…

I
due campioni
non sono
indipendenti
perchè
a coppie
sono
"collegati"

Un *contro*esem

Effetto di una sostanza omeopatica

Un campione casuale di individui di c
prima e **dopo** la somministrazione de

QUALE TEST POSSIAMO FARE?

Individuo	Pres. min. PRIMA	Pres. min. DOPO
A. Bianchini	83	75
B. Rossini	90	87
C. Verdini	66	62
E. Giallini	85	86
F. Turchini	72	70
…	…	…

I due campioni **non** sono **indipendenti** perchè a coppie sono "collegati"

Un altro controsempio

Altra situazione tipica: misura di una certa variabile effettuata su due unità statistiche legate tra loro: *campioni accoppiati*.

Statistica-4a-IntConf.pdf

Si vuole sapere se le donne tendono a sposare uomini che sono più alti di loro. Per questo si è scelto un campione di 50 coppie sposate, e si sono confrontate le altezze di moglie e marito calcolando la differenza di altezza: (h del marito – h della moglie). La differenza media è risultata pari a $\bar{x}_n = 11.2$ cm con una deviazione standard $s_n = 10.7$ cm.

RISOLVERE CON UN TEST!