



Corso di analisi dei dati sperimentali con R

Giorno 3 - Test statistici principali

Davide Gentilini

gentilini.davide@gmail.com

Antonino Oliveri

antonino.oliveri01@universitadipavia.it

Università degli Studi di Pavia

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2



Statistica inferenziale:

Insieme di metodi che consentono di generalizzare i risultati campionari ad una popolazione più ampia. La generalizzazione dei risultati dipende dal criterio di campionamento.

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Concetti preliminari

Ipotesi nulla e ipotesi alternativa



- 1 Il ricercatore formula una teoria (ipotesi) che descrive un fenomeno
 - Il fumo causa il tumore ai polmoni
 - I vaccini causano l'autismo
 - Il nuovo farmaco è più efficace di quello in circolazione
 - L'altezza media è cambiata nel tempo
- 2 Si conduce un esperimento per verificare l'ipotesi del ricercatore
 - Le conoscenze sul fenomeno fino a quel momento definiscono l'*ipotesi nulla* H_0 . L'ipotesi nulla si ritiene vera *fino a prova contraria*
 - Attraverso l'esperimento, il ricercatore **deve dimostrare** che le conoscenze sul fenomeno sono errate o incomplete e che la sua ipotesi (detta *alternativa*, H_1) chiarisce meglio il fenomeno in studio

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tabelle di contingenza $J \times K$

Tabelle di contingenza 2×2

Concetti preliminari

Verifica d'ipotesi



$$A) \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

$$B) \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

$$C) \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

A,C: test unidirezionali

B: test bidirezionale

Sotto H_0 : $W \rightarrow$ Statistica test con distribuzione nota

Esempio: Statistica di Wald

$$W = \frac{\hat{\theta} - \theta_0}{\sqrt{\text{Var}(\hat{\theta})}} \sim \mathcal{N}(0, 1)$$

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

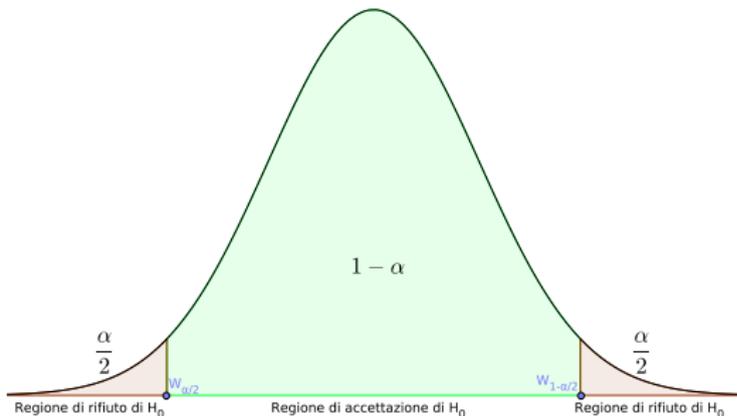
Test di associazione

Tabelle di contingenza $J \times K$

Tabelle di contingenza 2×2

Concetti preliminari

Rifiuto di H_0



$$W = \frac{\hat{\theta} - \theta_0}{\sqrt{\text{Var}(\hat{\theta})}} \sim \mathcal{N}(0, 1)$$

- $w_{\alpha/2}$, $w_{1-\alpha/2}$: valori critici della distribuzione del test, l'area al di fuori di questi due valori è pari a una quantità fissata a priori detta α

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tabelle di contingenza $J \times K$

Tabelle di contingenza 2×2

Caratteristiche dei test statistici

Realtà vs Test



Realtà	Test	
	Negativo	Positivo
H_0 vera	✓	✗
H_0 falsa	✗	✓

Risultato del test:

- Negativo: non rifiuta H_0
- Positivo: rifiuta H_0

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tablelle di contingenza $J \times K$

Tablelle di contingenza 2×2

Caratteristiche dei test statistici

Errore del primo tipo



Realtà	Test	
	Negativo	Positivo
H_0 vera	$1 - \alpha$	α
H_0 falsa	X	✓

α : livello di significatività, errore del primo tipo

- $\alpha = \Pr(\text{Rif. } H_0 | H_0)$
- $\alpha \rightarrow$ probabilità che il test restituisca un risultato falsamente positivo

$1 - \alpha$: livello fiduciario

- Indica il grado di fiducia attribuibile al risultato del test

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tabelle di contingenza $J \times K$

Tabelle di contingenza 2×2

Caratteristiche dei test statistici

Errore del secondo tipo



Realtà	Test	
	Negativo	Positivo
H_0 vera	$1 - \alpha$	α
H_0 falsa	β	$1 - \beta$

β : errore del secondo tipo

- $\beta = \Pr(\text{Non Rif. } H_0 | H_0^F)$
- $\beta \rightarrow$ probabilità che il test restituisca un risultato falsamente negativo

$1 - \beta$: potenza del test

- Capacità del test di rilevare la condizione espressa nell'ipotesi alternativa

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tabelle di contingenza $J \times K$

Tabelle di contingenza 2×2

Caratteristiche dei test statistici

α e β



Realtà	Test	
	Negativo	Positivo
H_0 vera	$1 - \alpha$	α
H_0 falsa	β	$1 - \beta$

- α e β sono inversamente proporzionali
- Solitamente si fissa α a priori e si cerca di minimizzare β
- I valori più utilizzati per α sono: 0.10, **0.05**, 0.01

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tablelle di contingenza $J \times K$

Tablelle di contingenza 2×2



P-value:

Misura di evidenza contro H_0 che accompagna il risultato di un test statistico. Permette di stabilire il risultato del test.

Indica:

- la probabilità che i dati siano compatibili con H_0 supponendo vera H_0
- la probabilità che l'eventuale discrepanza fra il test e H_0 sia dovuta al caso

Per convenzione, si rifiuta H_0 se $p < \alpha$.

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tablelle di contingenza $J \times K$

Tablelle di contingenza 2×2



Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tablelle di contingenza $J \times K$

Tablelle di contingenza 2×2

- "Accettare¹" H_0 non significa necessariamente che H_0 sia vera, significa che non si hanno evidenze sufficienti per rifiutarla
 - Uno studio con bassa potenza potrebbe non essere in grado di rilevare un effetto se questo esiste

¹Il termine corretto è "Non rifiutare"



Test statistici

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test statistici

Test parametrici e non parametrici



- Test parametrico: test in cui si suppone nota la distribuzione della caratteristica di interesse nella popolazione. Si fa inferenza su uno o più parametri della distribuzione. La conoscenza dei parametri implica la conoscenza totale della distribuzione
- Test non parametrico: test in cui non si fanno assunzioni sulla distribuzione della caratteristica. Più potente nel caso in cui l'assunto distribuzionale di un dato test parametrico non è valido

Concetti preliminari

Verifica d'ipotesi
Errore del primo tipo
Errore del secondo tipo
P-value

Test statistici

Test parametrici
Media
Differenza fra medie
Proporzione e differenza fra proporzioni
Correlazione
Test non parametrici
Media e differenza tra medie
Mediana e differenza fra mediane
Proporzione e differenza fra proporzioni
Correlazione
Test di associazione
Tabelle di contingenza $J \times K$
Tabelle di contingenza 2×2



Test parametrici

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

Test parametrici

- Media
- Differenza fra medie
- Proporzione e differenza fra proporzioni
- Correlazione

Test non parametrici

- Media e differenza tra medie
- Mediana e differenza fra mediane
- Proporzione e differenza fra proporzioni
- Correlazione
- Test di associazione
- Tablelle di contingenza $J \times K$
- Tablelle di contingenza 2×2

Test parametrici

Test parametrici sulla media



$$A) \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

$$B) \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

$$C) \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Sotto H_0 :

$$W = \frac{\bar{X} - \mu_0}{\sqrt{\text{Var}(\bar{X})}}$$

Se:

- $X \sim \mathcal{N}(\mu_0, \sigma_0^2)$, con μ_0 e σ_0 noti oppure se n è grande, W prende il nome di Z test e segue una distribuzione $\mathcal{N}(0, 1)$
- X è normale ma σ_0 non è noto, W prende il nome di T test e segue una distribuzione t_{n-1}

Concetti preliminari

Verifica d'ipotesi
Errore del primo tipo
Errore del secondo tipo
P-value

Test statistici

Test parametrici
Media
Differenza fra medie
Proporzione e differenza fra proporzioni
Correlazione
Test non parametrici
Media e differenza tra medie
Mediana e differenza fra mediane
Proporzione e differenza fra proporzioni
Correlazione
Test di associazione
Tabelle di contingenza $J \times K$
Tabelle di contingenza 2×2

Test parametrici

Test parametrici sulla media



Supponiamo che l'oggetto x sia un vettore che contiene il peso rilevato su un campione di soggetti sufficientemente grande. Vogliamo verificare se il peso medio della popolazione sia minore, diverso o maggiore di 80, secondo i casi A,B,C visti nella precedente diapositiva. Da studi precedenti si sa che $\sigma^2 = 160$.

```
library(BSDA)
```

```
#B
```

```
z.test(x,mu=80, sigma.x=sqrt(160))
```

One-sample z-Test

```
data: x
```

```
z = 1.1063, p-value = 0.2686
```

```
alternative hypothesis: true mean is not equal to 80
```

```
#A
```

```
z.test(x,mu=80, sigma.x=sqrt(160), alternative="less")
```

```
#C
```

```
z.test(x,mu=80, sigma.x=sqrt(160) alternative="greater")
```



Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media**
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
 - Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
 - Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test parametrici

Test parametrici sulla media



Supponiamo che l'oggetto x sia un vettore che contiene il peso rilevato su un campione di soggetti. Vogliamo verificare che il peso medio della popolazione sia pari a 80, l'ipotesi alternativa varierà secondo i casi A,B,C visti nella precedente diapositiva.

#B

```
t.test(x,mu=80)
```

#A

```
t.test(x,mu=80, alternative="less")
```

#C

```
t.test(x,mu=80, alternative="greater")
```

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media**
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
 - Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
 - Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test parametrici

Test parametrici sulla differenza fra medie



Per verificare se due campioni provengono dalla stessa popolazione

$$A) \begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$$

$$B) \begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

$$C) \begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{cases}$$

$$A) \begin{cases} H_0 : \delta = 0 \\ H_1 : \delta < 0 \end{cases}$$

$$B) \begin{cases} H_0 : \delta = 0 \\ H_1 : \delta \neq 0 \end{cases}$$

$$C) \begin{cases} H_0 : \delta = 0 \\ H_1 : \delta > 0 \end{cases}$$

$$\delta = \mu_1 - \mu_2$$

- $\delta < 0 \leftrightarrow \mu_1 < \mu_2$
- $\delta = 0 \leftrightarrow \mu_1 = \mu_2$
- $\delta > 0 \leftrightarrow \mu_1 > \mu_2$

Concetti preliminari

Verifica d'ipotesi
Errore del primo tipo
Errore del secondo tipo
P-value

Test statistici

Test parametrici
Media
Differenza fra medie
Proporzione e differenza fra proporzioni
Correlazione
Test non parametrici
Media e differenza tra medie
Mediana e differenza fra mediane
Proporzione e differenza fra proporzioni
Correlazione
Test di associazione
Tabelle di contingenza $J \times K$
Tabelle di contingenza 2×2

Test parametrici

Test parametrici sulla differenza fra medie



- Grandi campioni indipendenti e varianze note

`library(BSDA)`

`z.test(x=x1,y=x2,sigma.x=sigmax, sigma.y=sigmay)`

- Varianze ignote: si verifica che le varianze siano statisticamente uguali

`var.test(x1,x2)`

- Se il test non è significativo:
`t.test(x1,x2,var.equal=T)`
- Se il test è significativo:
`t.test(x1,x2)`

Se i campioni sono dipendenti si aggiunge l'opzione `paired=T` nel comando `t.test`

Concetti preliminari

Verifica d'ipotesi

Errore del primo tipo

Errore del secondo tipo

P-value

Test statistici

Test parametrici

Media

Differenza fra medie

Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici

Media e differenza tra medie

Mediana e differenza fra mediane

Proporzione e differenza fra proporzioni

Correlazione

Test di associazione

Tablelle di contingenza $J \times K$

Tablelle di contingenza 2×2

Test parametrici

Test parametrici sulla proporzione e sulla differenza fra proporzioni



Proporzione:

$$\begin{cases} H_0 : p = \pi_0 \\ H_1 : p \neq \pi_0 \end{cases}$$

Se X è dicotomica, segue una distribuzione binomiale, con media e varianza pari rispettivamente a:

$$\begin{aligned} \mathbb{E}[X] &= p \\ \text{Var}[X] &= \frac{p(1-p)}{n} \end{aligned}$$

Codificando X come variabile numerica 0/1, in \mathbb{R} è possibile stimare p tramite uno z-test (vedi slide 16).

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni**
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test parametrici

Test parametrici sulla proporzione e sulla differenza fra proporzioni



Nel 2010 il fornitore di un servizio ha stimato che il 70% dei clienti si riteneva soddisfatto. Si vuole verificare che, a distanza di 5 anni, tale proporzione sia aumentata. La società chiede una valutazione del servizio a 1000 utenti: 857 dichiarano di essere soddisfatti.

```
> table(sodd)
  0   1
143 857

> n<- sum(table(sodd))
> p0<- 0.7
> varx<- p0*(1-p0)/n
> z.test(sodd, mu=p0, sigma.x=sqrt(varx), alt="g")
```

```
z = 342.6, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0.7
```

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni**
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test parametrici

Test parametrici sulla proporzione e sulla differenza fra proporzioni



$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

Anche in questo caso, trattando le due variabili come vettori numerici 0/1, in R si può utilizzare uno z-test per stimare la differenza (vedi slide 19).

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni**
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test parametrici

Test parametrici sulla correlazione



$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

- Correlazione marginale:

```
library(psych)
corr.test(dati)
```

```
#Correzione confronti multipli
corr.test(dati, adjust="fdr")
```

```
#Intervalli di confidenza
print(corr.test(dati), short=F)
```

- Correlazione parziale:

```
library(ppcor)
pcor(dati)
```

Concetti preliminari

Verifica d'ipotesi
Errore del primo tipo
Errore del secondo tipo
P-value

Test statistici

Test parametrici
Media
Differenza fra medie
Proporzione e differenza fra proporzioni

Correlazione

Test non parametrici
Media e differenza tra medie
Mediana e differenza fra mediane
Proporzione e differenza fra proporzioni
Correlazione

Test di associazione
Tabelle di contingenza $J \times K$
Tabelle di contingenza 2×2



Test non parametrici

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione

Test non parametrici

- Media e differenza tra medie
- Mediana e differenza fra mediane
- Proporzione e differenza fra proporzioni
- Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test non parametrici

Test non parametrico sulla media e sulla differenza fra medie



Test di Wilcoxon: wilcox.test

#Media

```
> x<-rnorm(1000,mu=0)
> wilcox.test(x, mu=0)
```

Wilcoxon signed rank test with continuity correction

data: x

V = 2368, p-value = 0.5905

alternative hypothesis: true location is not equal to 0

#Differenza fra medie

```
> y<-rnorm(1000, mu=30)
> wilcox.test(x,y)
```

Wilcoxon rank sum test with continuity correction

data: x and y

W = 0, p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0

Concetti preliminari

Verifica d'ipotesi
Errore del primo tipo
Errore del secondo tipo
P-value

Test statistici

Test parametrici
Media
Differenza fra medie
Proporzione e differenza fra proporzioni
Correlazione

Test non parametrici
Media e differenza tra medie

Mediana e differenza fra mediane
Proporzione e differenza fra proporzioni
Correlazione
Test di associazione
Tabelle di contingenza $J \times K$
Tabelle di contingenza 2×2

Test non parametrici

Test non parametrico sulla mediana



Sign-test:

```
> library(BSDA)

> median0<-3

> SIGN.test(dati$x, md = median0)
```

One-sample Sign-Test

s = 7, p-value = 0.07031

alternative hypothesis: true median is not equal to 3

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane**
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test non parametrici

Test non parametrico sulla differenza fra mediane



Test di Kruskal-Wallis:

```
kruskal.test(outcome ~ group)
```

```
#NB: la classe della variabile "group" deve essere di tipo "factor"
```

```
>data("airquality")
```

```
#data serve solo a caricare un dataset già presente in R
```

```
>kruskal.test(Ozone ~ Month, data = airquality)
```

Kruskal-Wallis rank sum test

```
data: Ozone by Month
```

```
Kruskal-Wallis chi-squared = 29.267, df = 4, p-value = 6.901e-06
```

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane**
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tablette di contingenza $J \times K$
 - Tablette di contingenza 2×2

Test non parametrici

Test non parametrico sulla proporzione e sulla differenza fra proporzioni



Per entrambi i tipi di test in R si può usare il comando *prop.test*

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni**
 - Correlazione
- Test di associazione
 - Tablelle di contingenza $J \times K$
 - Tablelle di contingenza 2×2

Test non parametrici

Test non parametrici sulla proporzione e sulla differenza fra proporzioni



Nel 2010 il fornitore di un servizio ha stimato che il 70% dei clienti si riteneva soddisfatto. Si vuole verificare che, a distanza di 5 anni, tale proporzione sia aumentata. La società chiede una valutazione del servizio a 1000 utenti: 857 dichiarano di essere soddisfatti.

```
> table(sodd)
1_si 2_no
857  143
> prop.test(table(sodd),p=0.70, alt="g")
```

```
data: table(sodd), null probability 0.7
X-squared = 116.63, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is greater than 0.7
```

prop.test considera come modalità di successo la prima che viene restituita dal comando *table*

Concetti

preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test non parametrici

Test non parametrici sulla proporzione e sulla differenza fra proporzioni



Si vuole valutare se all'interno di una grande azienda la proporzione di dipendenti che ha ricevuto una promozione sia diversa fra uomini e donne

```
> table(azienda$genere, azienda$promozione)
```

```
      1_SI 2_NO  
1_M    48  202  
2_F    27  123
```

```
> prop.test(table(azienda$genere, azienda$promozione))
```

```
X-squared = 0.02735, df = 1, p-value = 0.8686  
alternative hypothesis: two.sided  
95 percent confidence interval:  
-0.07184314  0.09584314  
sample estimates:  
prop 1 prop 2  
0.192  0.180
```

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni**
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2

Test non parametrici

Test non parametrici sulla correlazione



Per ottenere una versione non parametrica dei test visti nella slide 23 basta aggiungere l'opzione *method="Spearman"*

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione**
- Test di associazione
 - Tabelle di contingenza $J \times K$
 - Tabelle di contingenza 2×2



Test di associazione

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione**
 - Tablelle di contingenza $J \times K$
 - Tablelle di contingenza 2×2

Test di associazione

Test per tabelle di contingenza $J \times K$



$$\begin{cases} H_0 : \text{Assenza di associazione} \\ H_1 : \text{Presenza di associazione} \end{cases}$$

Per tabelle con numero di righe e colonne qualsiasi si può utilizzare il comando *chisq.test*

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$**
 - Tabelle di contingenza 2×2

Test di associazione

Test per tabelle di contingenza $J \times K$



Si vuole indagare se vi è un'associazione fra l'età e il numero medio di ore trascorse giornalmente sui social network. Si hanno a disposizione dei dati in forma tabulare.

```
> sn
      [10,30) [30,50) 50+
Meno di 1 ora      35      71 101
Fra 1 e 2 ore      95      68  28
Fra 2 e 3 ore      60      15   6
Più di 3 ore      10       6   5
```

```
> chisq.test(sn)
```

Pearson's Chi-squared test

```
data: sn
```

```
X-squared = 118.23, df = 6, p-value < 2.2e-16
```

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test non parametrici
 - Media e differenza tra medie
 - Mediana e differenza fra mediane
 - Proporzione e differenza fra proporzioni
 - Correlazione
- Test di associazione
 - Tabelle di contingenza $J \times K$**
 - Tabelle di contingenza 2×2

Test di associazione

Test per tabelle di contingenza 2×2



- Test Chi-quadrato di Pearson
 - *chisq.test*
- Test esatto di Fisher
 - Adatto per campioni con $n \leq 40$, *fisher.test*
- OR
 - Adatto generalmente in qualsiasi contesto, in particolare negli studi caso-controllo. $H_0 : OR = 1$
- RR
 - Adatto negli studi prospettici. $H_0 : RR = 1$

Concetti preliminari

Verifica d'ipotesi
Errore del primo tipo
Errore del secondo tipo
P-value

Test statistici

Test parametrici
Media
Differenza fra medie
Proporzione e differenza fra proporzioni
Correlazione
Test non parametrici
Media e differenza tra medie
Mediana e differenza fra mediane
Proporzione e differenza fra proporzioni
Correlazione
Test di associazione
Tabelle di contingenza $J \times K$
Tabelle di contingenza 2×2



Test di associazione

Test per tabelle di contingenza 2x2

Si vuole valutare se una certa esposizione risulta essere un fattore di rischio per un dato evento.

```
> table(epid$esposizione, epid$status)
```

	evento	non_evento
esposto	146	57
non_esposto	43	232

```
> library(epiR)
```

```
> epi.2by2(table(epid$esposizione, epid$status))
```

	Outcome +	Outcome -	Total	Inc risk	Odds
Exposed +	146	57	203	71.9	2.561
Exposed -	43	232	275	15.6	0.185
Total	189	289	478	39.5	0.654

Point estimates and 95 % CIs:

Inc risk ratio (RR) 4.60 (3.45, 6.13)

Odds ratio (OR) 13.82 (8.84, 21.60)

X2 test statistic: 154.767 p-value: <0.001

Concetti preliminari

- Verifica d'ipotesi
- Errore del primo tipo
- Errore del secondo tipo
- P-value

Test statistici

- Test parametrici
 - Media
 - Differenza fra medie
 - Proporzione e differenza fra proporzioni
 - Correlazione

Test non parametrici

- Media e differenza tra medie
- Mediana e differenza fra mediane
- Proporzione e differenza fra proporzioni
- Correlazione

Test di associazione

- Tabelle di contingenza JxK
- Tabelle di contingenza 2x2