

METODI STATISTICI PER LA RICERCA SOCIALE L'ANALISI DELL'ASSOCIAZIONE TRA VARIABILI CATEGORIALI IN

Alessandra Mattei

Dipartimento di Statistica, Informatica, Applicazioni (DiSIA)
Università degli Studi di Firenze
mattei@disia.unifi.it

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Analisi dell'associazione tra due caratteri

Curare la tosse

- L'uso del miele per curare la tosse notturna e migliorare la qualità del sonno in bambini con infezioni del tratto respiratorio superiore
- Variabili
 - Gender = Genere (0 = Femmina; 1 = Maschio)
 - Age = Età (in anni)
 - treatment = Trattamento
(A = Destrometorfano (antitossivo) al gusto di miele;
B = Miele; C = No trattamento)
 - cba.cat = La tosse disturba il sonno
(1 = Per niente; 2 = Poco; 3 = Abbastanza; 4 = Molto)

Analisi dell'associazione tra due caratteri

```
> rm(list=ls())
>
> ##Dati: Dati non reali
> load("Dati/Honey3ArmTrials.RData")
> ls()
[1] "Honey"
>
> str(Honey)
'data.frame': 210 obs. of 2 variables:
 $ treatment: Factor w/ 3 levels "A","B","C": 1 1 3 2 1 2 3 2 3 1 ..
 $ cba.cat   : num  2 2 3 4 1 2 3 1 2 4 ...
>
>
> attach(Honey)
```

Distribuzioni di frequenza

```
> ##Distribuzioni di frequenza assolute
> ##Marginali
> table(treatment)
treatment
A  B  C
66 70 74
> table(cba.cat)
cba.cat
1  2  3  4
48 70 42 50
>
> ##Congiunte
> table(treatment, cba.cat)
      cba.cat
treatment 1  2  3  4
A  12 26 10 18
B  22 26 10 12
C  14 18 22 20
```

Distribuzioni di frequenza

```
> ##Distribuzioni di frequenza relative
> nr<- nrow(Honey)
> nr
[1] 210
>
> #Marginali
> table(treatment)/nr
treatment
A          B          C
0.3142857 0.3333333 0.3523810
> table(cba.cat)/nr
cba.cat
1          2          3          4
0.2285714 0.3333333 0.2000000 0.2380952
>
> ##Congiunte
> table(treatment, cba.cat)/nr
      cba.cat
treatment  1          2          3          4
A 0.05714286 0.12380952 0.04761905 0.08571429
B 0.10476190 0.12380952 0.04761905 0.05714286
C 0.06666667 0.08571429 0.10476190 0.09523810
```

Distribuzioni di frequenza

```
> ##Relative condizionate cba.cat|treatment
> table(treatment)
treatment
  A  B  C
66 70 74
>
> table(cba.cat[treatment=="A"])/sum(treatment=="A")
      1      2      3      4
0.1818182 0.3939394 0.1515152 0.2727273
>
> table(cba.cat[treatment=="B"])/sum(treatment=="B")
      1      2      3      4
0.3142857 0.3714286 0.1428571 0.1714286
>
> table(cba.cat[treatment=="C"])/sum(treatment=="C")
      1      2      3      4
0.1891892 0.2432432 0.2972973 0.2702703
```

Test χ^2 di indipendenza

```
> #Test chi-quadrato di indipendenza  
> Xsq<-chisq.test(table(treatment, cba.cat), correct=FALSE)  
> Xsq
```

Pearson's Chi-squared test

```
data:  table(treatment, cba.cat)  
X-squared = 13.599, df = 6, p-value = 0.03445
```

Test χ^2 di indipendenza

```
> Xsq$observed # tabella di frequenza
```

```
  cba.cat
```

```
treatment  1  2  3  4
```

```
  A 12 26 10 18
```

```
  B 22 26 10 12
```

```
  C 14 18 22 20
```

```
>
```

```
> Xsq$expected # Frequenze attese
```

```
  cba.cat
```

```
treatment      1      2      3      4
```

```
  A 15.08571 22.00000 13.2 15.71429
```

```
  B 16.00000 23.33333 14.0 16.66667
```

```
  C 16.91429 24.66667 14.8 17.61905
```

```
>
```

```
> Xsq$observed - Xsq$expected # Residui
```

```
  cba.cat
```

```
treatment      1      2      3      4
```

```
  A -3.085714  4.000000 -3.200000  2.285714
```

```
  B  6.000000  2.666667 -4.000000 -4.666667
```

```
  C -2.914286 -6.666667  7.200000  2.380952
```


Studio dell'associazione in tabelle 2×2

Corso sull'auto-palpazione del seno e pratica dell'auto-palpazione del seno

- Associazione tra la frequenza di un corso di formazione sull'auto-palpazione del seno e la pratica dell'auto-palpazione del seno
- Variabili
 - Z = Frequenza del corso (0 = No; 1 = Si)
 - bse = Pratica dell'auto-palpazione del seno a un anno dal corso

Studio dell'associazione in tabelle 2×2

```
> #####
> #Studio dell'associazione in tabelle 2X2
> #####
> rm(list=ls())
> ##Dati
> Bse <- read.table("Dati/BseFake.txt", header=T)
>
> str(Bse)
'data.frame': 800 obs. of 2 variables:
 $ course: int  1 0 0 1 1 1 1 1 0 1 ...
 $ bse : int  1 1 0 1 1 1 1 1 1 1 ...
>
> attach(Bse)
>
> table(course,bse)
      bse
course  0  1
      0 87 335
      1 47 331
```

Studio dell'associazione in tabelle 2×2

```
> ##Sintesi campionarie
> n0<- sum(1-course)
> n0
[1] 422
>
> bse0 <-sum(bse[course==0])
> bse0
[1] 335
>
> n1<- sum(course)
> n1
[1] 378
>
> bse1<-sum(bse[course==1])
> bse1
[1] 331
```

Studio dell'associazione in tabelle 2×2

```
> ##H0: p1-p0 = 0 versus Ha: p1-p0 != 0  
> prop.test(c(bse1, bse0), c(n1,n0), correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data:  c(bse1, bse0) out of c(n1, n0)  
X-squared = 9.5733, df = 1, p-value = 0.001974  
alternative hypothesis: two.sided  
95 percent confidence interval:  
0.0308689 0.1327761  
sample estimates:  
prop 1      prop 2  
0.8756614 0.7938389  
>  
> #Test chi-quadrato di indipendenza  
> Xsq<-chisq.test(table(course, bse), correct=FALSE)  
> Xsq
```

Pearson's Chi-squared test

```
data:  table(course, bse)  
X-squared = 9.5733, df = 1, p-value = 0.001974
```

Studio dell'associazione in tabelle 2×2

```
> ##Misure di associazione in tabelle 2x2"
> ##Rischio relativo e odds-ratio
> #install.packages("epitools")
> library(epitools)
> ##Rischio relativo
> rr<-riskratio(course, bse)
> rr$measure
```

risk ratio with 95% C.I.

Predictor	estimate	lower	upper
0	1.000000	NA	NA
1	1.103072	1.037068	1.173277

Studio dell'associazione in tabelle 2×2

```
> ##Odds-ratio
> or<-oddsratio(course, bse)
> or$measure
odds ratio with 95% C.I.
Predictor estimate      lower      upper
      0 1.000000         NA         NA
      1 1.825112 1.245411 2.70348
>
> or.ic90<-oddsratio(course, bse, conf.level=0.90)
> or.ic90$measure
odds ratio with 90% C.I.
Predictor estimate      lower      upper
      0 1.000000         NA         NA
      1 1.825112 1.323657 2.535799
>
> detach(Bse)
```