

# METODI STATISTICI PER LA RICERCA SOCIALE

## CAPITOLO 7. CONFRONTO TRA DUE GRUPPI

### ESERCITAZIONE

Alessandra Mattei

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

LM-88 SOCIOLOGIA E RICERCA SOCIALE

# Esercizio 1

Interessa confrontare il prezzo al litro (in euro) dell'olio extra vergine di oliva prodotto in Toscana e in Liguria. Il prezzo al litro (in euro) dell'olio extra vergine di oliva prodotto da un campione di 4 aziende Toscane è (7.2, 9.5, 7.0, 7.5). Il prezzo al litro (in euro) dell'olio extra vergine di oliva prodotto da un campione di 7 aziende della Liguria è risultato in media di 8.5 Euro con varianza campionaria 1.44. Rispondere ai seguenti quesiti supponendo che le popolazioni siano normali e abbiano uguale varianza.

- (a) Calcolare una stima della varianza  $\sigma^2$
- (b) Stabilire con un test al livello di significatività del 10% se il prezzo dell'olio extra vergine di oliva prodotto in Toscana e in Liguria è diverso. (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 10%; (iii) Calcolare il valore della statistica test e prendere una decisione
- (c) In base al risultato ottenuto al punto (b) stabilire se il  $p$ -valore per il test al punto (b) sarebbe stato minore/uguale/maggiore di 0.10 (Giustificare la risposta).
- (d) Costruire un intervallo di confidenza al 90% per la differenza tra le medie del prezzo al litro (in euro) dell'olio extra vergine di oliva prodotto in Toscana e in Liguria. Confrontare il risultato dell'intervallo di confidenza con il risultato del test ottenuto al punto precedente.

## Esercizio 2

Per verificare l'efficacia di un corso di recupero per adolescenti con problemi in matematica, gli organizzatori hanno deciso di sottoporre 5 studenti scelti a caso a due prove successive: la prima all'inizio del corso, e la seconda alla fine. I punteggi (in centesimi) ottenuti nelle prove sono riportati in tabella

| Soggetto | 1  | 2  | 3  | 4  | 5  |
|----------|----|----|----|----|----|
| Prima    | 60 | 65 | 50 | 75 | 48 |
| Dopo     | 68 | 80 | 40 | 90 | 78 |

Si assuma che le variabili aleatorie dei punteggi si distribuiscano secondo una distribuzione Normale.

- (a) Stimare la differenza tra le medie e l'errore standard dello stimatore della differenza tra le medie
- (b) Sottoporre a test l'ipotesi nulla "il corso non ha modificato le capacità dei frequentanti": (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 5%; (iii) Calcolare il valore della statistica test e prendere una decisione
- (c) Costruire un intervallo di confidenza al livello di confidenza  $1 - \alpha = 0.95$  per la differenza tra le medie. Confrontare il risultato dell'intervallo di confidenza con il risultato del test ottenuto al punto precedente.

## Esercizio 3

La seguente tabella mostra la distribuzione di frequenza congiunta del tipo di trattamento e della qualità della vita per un gruppo di soggetti in dialisi.

| Trattamento          | Qualità della vita |              |
|----------------------|--------------------|--------------|
|                      | Medio - Bassa      | Medio - Alta |
| Nuovo trattamento    | 320                | 680          |
| Trattamento standard | 456                | 744          |

- (a) Calcolare la proporzione di pazienti con qualità della vita medio-alta per pazienti esposti al nuovo trattamento e per pazienti esposti al trattamento standard
- (b) Verificare l'ipotesi che la proporzione di pazienti con qualità della vita medio-alta sia maggiore tra pazienti esposti al nuovo trattamento rispetto a quella di pazienti esposti al trattamento standard: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 5%; (iii) Calcolare il valore della statistica test e prendere una decisione
- (c) Calcolare il  $p$ -valore per il test al punto (b) e interpretarlo.
- (d) Costruire un intervallo di confidenza al livello di confidenza del 95% per la differenza tra proporzioni nella popolazione.
- (e) Calcolare gli odds a favore della qualità della vita medio-alta per pazienti esposti al nuovo trattamento e per pazienti esposti al trattamento standard. Calcolare quindi il rapporto degli odds
- (f) Costruire un intervallo di confidenza al livello di confidenza del 95% per il rapporto degli odds calcolato al punto (e)

## Esercizio 4

La seguente tabella mostra la distribuzione di frequenza congiunta ottenuta classificando un campione di cittadini di un certo comune rispetto al livello di istruzione e al livello di soddisfazione relativa ai servizi sociali presenti sul territorio

| Livello di istruzione | Livello di Soddisfazione |             |                   |
|-----------------------|--------------------------|-------------|-------------------|
|                       | Insoddisfatto            | Soddisfatto | Molto Soddisfatto |
| No Diploma            | 18                       | 60          | 40                |
| Diploma o più         | 48                       | 72          | 40                |

- (a) Costruire la distribuzione di frequenza relativa condizionata del livello di soddisfazione dato un livello di istruzione pari al diploma o più elevato
- (b) Calcolare la frequenza teorica di indipendenza per cittadini senza diploma e livello di soddisfazione pari a molto soddisfatto. Calcolare il corrispondente residuo (ossia la corrispondente contingenza)
- (c) Verificare l'ipotesi che livello di soddisfazione e livello di istruzione siano indipendenti. (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 10%; (iii) Prendere una decisione sapendo che il valore della statistica test  $\chi^2$  è  $\chi^{2,oss} = 9.25$

## Esercizio 5

La seguente tabella mostra le distribuzioni di frequenza marginali delle preferenze per il tempo libero e del titolo di studio su un campione di giovani adulti

- (a) Costruire la distribuzione di frequenza assoluta congiunta la tabella che segue ipotizzando che preferenze per il tempo libero e titolo di studio siano variabili statisticamente indipendenti

| Titolo di studio            | Cinema e TV | Lettura | Musica | Sport | Totale |
|-----------------------------|-------------|---------|--------|-------|--------|
| Scuola dell'Obbligo         |             |         |        |       | 360    |
| Diploma                     |             |         |        |       | 720    |
| Diploma universitario o più |             |         |        |       | 540    |
| Totale                      | 486         | 162     | 324    | 648   | 1620   |

- (b) Supporre di fare un test  $\chi^2$  per valutare l'ipotesi che preferenze per il tempo libero e titolo di studio siano statisticamente indipendenti. Specificare la distribuzione (asintotica) della statistica test.
- (c) Scrivere la regione critica del test al punto (b) per un livello di significatività  $\alpha = 0.01$ .