

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Compito di
Metodi Statistici per la Ricerca Sociale
14 Giugno 2013

Esercizi

Esercizio 1.

Un campione di donne è stato classificato secondo la condizione lavorativa e la zona geografica di residenza come segue:

| Zona residenza | Condizione lavorativa | |
|-------------------|-----------------------|----------|
| | Non Occupata | occupata |
| Nord | 150 | 350 |
| Centro-Sud | 330 | 270 |

1. Costruire le distribuzioni condizionate relative della condizione lavorativa data la zona di residenza
2. Verificare se condizione lavorativa e zona di residenza sono statisticamente indipendenti al livello di significatività del 5%.
3. Calcolare l'odds ratio considerando come probabilità di successo la probabilità che la donna sia occupata. Costruire quindi un intervallo di confidenza al livello di confidenza del 95%. Interpretare l'odds ratio e il relativo intervallo di confidenza

Esercizio 2.

Su un campione di soggetti si sono rilevati i seguenti dati relativi al numero di anni di istruzione e il reddito annuo (in migliaia di Euro):

| | | | | | |
|----------------------------|----|----|----|----|----|
| Anni di istruzione (X) | 8 | 13 | 13 | 18 | 20 |
| Reddito (Y) | 15 | 30 | 35 | 50 | 70 |

1. Stimare i coefficienti di regressione del modello di regressione $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ che pone il reddito annuo (variabile Y) in funzione degli anni di istruzione (variabile X).
2. Stimare la varianza degli errori, σ^2 .
3. Sottoporre a test il sistema di ipotesi $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ al livello di significatività del 1%

Esercizio 3.

In uno studio sull'abuso di alcool nelle discoteche è stato rilevato il tasso alcolemico su un campione di 104 di soggetti (maschi) all'uscita da discoteche. I 104 soggetti sono classificati per zona di residenza: Nord, Centro, Sud, Isole. Di seguito i risultati di alcune analisi

| Fonte di variabilità | Somma dei Quadrati | <i>GdL</i> | Media dei Quadrati |
|----------------------|--------------------|------------|--------------------|
| Tra gruppi | 2.54 | | |
| Entro gruppi | | | |
| Totale | 3.56 | | |

1. Completare la tavola di analisi della varianza
2. **SOLO PER ESAME DA 9 CFU:** Verificare l'ipotesi che il tasso alcolemico medio sia lo stesso nelle diverse zone di residenza al livello di significatività del 1%.

Domande Teoriche

Domanda 1.

In uno studio finalizzato a valutare se esiste una differenza tra il prezzo medio delle case a Milano e a Roma sono stati rilevati i prezzi di un campione di case nelle due città. Nell'ipotesi che le popolazioni siano normali con uguale varianza, l'intervallo di confidenza al livello di confidenza $1 - \alpha = 0.98$ per la differenza tra il prezzo medio delle case a Milano e a Roma, μ_1 e μ_2 , è risultato $(-2\,000; 5\,000)$. Interpretare il risultato. Se fosse effettuato un test di ipotesi sulla differenza tra le due medie, il valore del p -value sarebbe maggiore, minore o uguale a 0.02? (Giustificare la risposta)

Domanda 2.

Si consideri un modello di regressione lineare multipla $Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \epsilon_i$. Cosa si intende per ipotesi di omoschedasticità?

Domanda 3. (SOLO PER ESAME DA 9 CFU)

Si supponga di voler studiare l'associazione tra tasso di disoccupazione, PIL e collocazione geografica nei paesi dell'Europa. Considerando una classificazione geografica in 3 zone (Nord, Centro, Sud), scrivere l'equazione del modello di regressione che pone il tasso di disoccupazione in funzione del PIL e della collocazione geografica ipotizzando che esista interazione tra PIL e collocazione geografica (prendere come zona di riferimento il Sud). In tale modello cosa rappresenta il coefficiente di regressione parziale relativo al PIL?