

Scuola di Scienze Politiche
Corso di Laurea LM–88 SOCIOLOGIA E RICERCA SOCIALE
Compito di Metodi Statistici per la Ricerca Sociale
13 Febbraio 2017

Parte applicativa

Esercizio 1

In uno studio finalizzato a valutare se l'organizzazione e la partecipazione a corsi di formazione possa aiutare giovani in cerca di prima occupazione a trovare un'occupazione si sono ottenuti i seguenti dati relativi alla partecipazione a un corso di formazione di durata 9 mesi e allo status occupazionale a 6 mesi dalla fine del corso di un campione di soggetti in cerca di prima occupazione:

Partecipazione al corso di formazione	Status occupazionale	
	Non occupato	Occupato
No	225	275
Si	150	350

- (a) Stimare la proporzione di soggetti occupati tra coloro che hanno partecipato al corso di formazione e coloro che non hanno partecipato al corso di formazione
- (b) Costruire un intervallo di confidenza al livello di confidenza del 97% per la differenza tra la proporzione di soggetti occupati tra coloro che hanno partecipato al corso di formazione e la proporzione di soggetti occupati tra coloro che non hanno partecipato al corso di formazione
- (c) Verificare l'ipotesi che la proporzione di soggetti occupati tra coloro che hanno partecipato al corso di formazione sia maggiore della proporzione di soggetti occupati tra coloro che non hanno partecipato al corso di formazione. (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) scrivere la regione critica (di rifiuto) al livello di significatività del 3%; (iii) calcolare il valore osservato della statistica test e prendere una decisione.
- (d) Calcolare il p -valore per il test al punto (c) e interpretare il valore che si ottiene confrontandolo anche con la decisione presa al punto (c)

Esercizio 2

Una ricerca sull'infanzia ha rilevato l'età (in mesi), variabile X , in cui 21 bambini hanno cominciato a parlare e il punteggio ottenuto dagli stessi soggetti, una volta diventati adulti, in un test attitudinale, variabile Y . Di seguito alcune statistiche descrittive

$$\sum_{i=1}^{21} x_i = 294 \quad \sum_{i=1}^{21} (x_i - \bar{x})^2 = 1262 \quad \sum_{i=1}^{21} y_i = 1965.6 \quad \sum_{i=1}^{21} (y_i - \bar{y})^2 = 3912 \quad \sum_{i=1}^{21} (x_i - \bar{x}) \cdot (y_i - \bar{y}) = -1426.06$$

- (a) Stimare il coefficiente di correlazione tra l'età in cui i bambini hanno cominciato a parlare e il punteggio ottenuto al test attitudinale
- (b) Determinare le stime dei minimi quadrati dei coefficienti della retta di regressione che pone il risultato al test attitudinale (Y) in funzione dell'età in cui i bambini hanno cominciato a parlare (X): $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- (c) Stimare la varianza degli errori
- (d) Sottoporre a test l'ipotesi $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$ al livello di significatività del 5%: (i) scrivere la regione critica (di rifiuto) al livello di significatività del 5%; (ii) calcolare il valore osservato della statistica test e prendere una decisione

Esercizio 3

Su un campione di $n = 34$ soggetti disoccupati in cerca di nuova occupazione si sono rilevate le seguenti variabili: Y = livello di depressione (misurato con un opportuno indice); X_1 = Titolo di studio (1 = Licenza media inferiore o meno; 2 = Diploma; 3 = Laurea o più); X_2 = Stato civile (1 = Single; 2 = Sposato/a; 3 = Separato/Divorziato; 4 = Vedovo/a). Si supponga di adattare un modello di regressione che pone il livello di depressione, Y in funzione delle variabili esplicative X_1 e X_2 . Si consideri tale modello come modello esteso. Per tale modello la somma dei quadrati dei residui è $SSE_e = 425.73$. Si supponga di voler confrontare tale modello con un modello ridotto in cui si considera come variabile esplicativa solo il titolo di studio, sapendo che per il modello ridotto la somma dei quadrati dei residui è $SSE_r = 490.80$.

- Scrivere l'ipotesi nulla e l'ipotesi alternativa relativa al confronto dei due modelli. Scrivere quindi la regione critica al livello di significatività $\alpha = 0.10$
- Calcolare il valore osservato della statistica test e prendere una decisione.
- Sulla base del risultato ottenuto al punto (b), stabilire se il p -valore per tale test è minore, maggiore o uguale a 0.10 (Giustificare la risposta).

Domande di Teoria

Domanda 1

Su un campione di comuni di un certo paese le variabili Y = tasso di criminalità (numero di episodi per 1000 residenti), X_1 = reddito mediano (in migliaia di euro), e X_2 = area in cui si trova il comune (1 = montana; 2 = rurale; 3 = urbana). La tabella sottostante mostra i risultati dell'adattamento di un modello di regressione in cui il tasso di criminalità è posto in funzione del reddito mediano (scartato dalla media pari a 35 000 euro), dell'area in cui si trova il comune e della loro interazione.

Variabile	Coefficiente	Errore Standard
Costante	5.26	1.63
Reddito mediano ($X_1 - 35$)	-0.80	0.25
Area (Montana) (X_2)		
Rurale	0.34	0.28
Urbana	1.22	0.31
Interazione ($(X_1 - 35) \cdot X_2$)		
Reddito · Rurale	-0.21	0.08
Reddito · Urbana	-0.42	0.11

- Scrivere l'equazione di regressione stimata
- Interpretare il coefficiente relativo al reddito, il coefficiente relativo all'area urbana e il coefficiente relativo all'interazione tra reddito e area urbana
- Calcolare il tasso di criminalità atteso per tre comuni, A, B, e C, aventi reddito mediano pari a 37, 000 ($X_1 = 37$) e localizzati rispettivamente in un'area montana (comune A); rurale (comune B) e urbana (comune C)

Domanda 2

Su un campione di disoccupati in cerca di nuova occupazione che hanno frequentato dei corsi di formazione si sono rilevate le variabili Y = status occupazione a sei mesi dal termine del corso di formazione (0 = Disoccupato; 1 = occupato) e X = durata (in giorni) del corso di formazione. La seguente tabella mostra i risultati dell'adattamento di modello di regressione logistica:

Variabile	Coefficiente	Errore standard
Costante	-1.789	0.332
Durate del corso	0.015	0.003

- Scrivere l'equazione che definisce la provabilità stimata $\hat{\pi}_i = \hat{Pr}(Y_i = 1|X_i = x_i)$.
- Interpretare il coefficiente relativo alla durata del corso in termini di odds.
- Calcolare la probabilità stimata di lavorare per un soggetto che ha frequentato un corso di formazione di 90 giorni