

Esercitazione di Metodi Statistici per la Ricerca Sociale
13 dicembre 2017

Esercizio 1

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

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

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 è $\chi^{2,oss} = 9.25$.

Esercizio 2

In uno studio sui comportamenti di giovani adulti, su un campione di $n = 28$ soggetti di età compresa tra i 20 e i 40 anni si sono rilevate le seguenti variabili $Y =$ Numero di libri letti in un anno; $X_1 =$ età e $X_2 =$ Titolo di Studio (1 = Obbligo o meno, 2 = Diploma, 3 = Laurea o più). I seguenti risultati sono relativi alle stime dei minimi quadrati di un modello di regressione che pone il numero di libri letti in un anno, Y , in funzione delle variabili esplicative età (scartata dalla media pari a 30 anni), e Titolo di Studio (per cui è considerata come categoria di riferimento la scuola dell'obbligo o meno):

Variabile	Coeff.	E.S.
Costante	1.92	0.37
Età - 30	0.16	0.05
Livello di istruzione		
Diploma	1.39	0.28
Laurea o più	1.32	0.36
$SST = 136.96$		$SSE = 36.61$

- (a) Scrivere l'equazione di regressione stimata
- (b) Stimare il numero medio di libri letti da un soggetto di 35 anni laureato
- (c) Stimare la devianza del modello
- (d) Confrontare il modello stimato con il modello nullo. (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 1%; (iii) Calcolare il valore della statistica test e prendere una decisione

Esercizio 3

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) 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.
- (b) Calcolare il p-valore per il test al punto (a) e interpretare il valore che si ottiene confrontandolo anche con la decisione presa al punto (a).

Esercizio 4

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

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

Esercizio 5

La seguente tabella riporta per un campione di 5 studenti il punteggio (misurato su scala 0-100) a un test finalizzato a valutare le conoscenze di matematica di matricole di una certa scuola prima e dopo un corso di introduzione alla matematica

Studente	Punteggio al test				
	1	2	3	4	5
Prima	66	62	67	68	75
Dopo	72	62	74	77	83

Si supponga che il punteggio al test prima e dopo il corso abbia distribuzione Normale con medie μ_{pre} e μ_{post} e varianze σ_{pre}^2 e σ_{post}^2 .

(a) Stimare la media e la varianza delle differenze tra il punteggio al test dopo il corso e il punteggio al test prima del corso

Si supponga di voler valutare l'ipotesi che la media del punteggio al test prima del corso sia inferiore alla media del punteggio al test dopo il corso (ipotesi alternativa)

(b) (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 5%

(c) Calcolare il valore della statistica test e prendere una decisione

(d) Il p-valore per tale test è $p = 0.018$. Interpretare tale valore anche sulla base del risultato ottenuto al punto (c)

Esercizio 6

La seguente tabella mostra il numero di anni di studio della lingua inglese e il punteggio ottenuto in un test di conoscenza della lingua per un campione di 6 persone

Anni di studio (X)	6	8	7	4	12	5
Punteggio (Y)	74	78	79	56	89	71

(a) Determinare le stime dei minimi quadrati dei coefficienti della retta di regressione che pone il punteggio (Y) in funzione del numero di anni di studio (X): $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

(b) Stimare la varianza degli errori

(c) Si consideri il test delle ipotesi $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$: (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

(d) Stimare il punteggio medio per soggetti con 10 anni di studio e costruire il relativo intervallo di confidenza al livello di confidenza del 95%.