

Scuola di Scienze Politiche

Corso di Laurea LM-88 in SOCIOLOGIA E RICERCA SOCIALE

Compito di Metodi Statistici per la Ricerca Sociale

06 Luglio 2017

Parte applicativa

Esercizio 1

In un certo paese viene realizzato un programma sociale volto a contrastare la povertà a medio e lungo termine. Famiglie beneficiarie del programma ricevono periodicamente dei contributi in denaro. La variabile di principale interesse è una variabile binaria uguale a 1 se tutti i componenti della famiglia in età scolare frequentano regolarmente la scuola e zero altrimenti

Famiglia beneficiaria	Frequenza regolare della scuola	
	No	Si
No	300	1200
Si	180	820

- (a) (i) Calcolare la proporzione di famiglie in cui tutti i componenti della famiglia in età scolare frequentano regolarmente la scuola tra famiglie beneficiarie e famiglie non beneficiarie. (ii) Calcolare gli odds a favore dell'evento "tutti i componenti della famiglia in età scolare frequentano regolarmente la scuola" tra le famiglie beneficiarie e tra le famiglie non beneficiarie.
- (b) (i) Calcolare il rapporto degli odds (odds ratio). (ii) Costruire un intervallo di confidenza per l'odd ratio al livello di confidenza del 95%.

Interessa verificare l'ipotesi che la proporzione di famiglie in cui tutti i componenti della famiglia in età scolare frequentano regolarmente la scuola sia superiore tra famiglie beneficiarie che tra famiglie non beneficiarie

- (c) (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 5%;
- (d) (i) Calcolare il valore della statistica test e prendere una decisione; (ii) Calcolare il p -valore. Interpretare tale valore anche sulla base del risultato ottenuto al punto (i)

Esercizio 2

Le seguenti statistiche descrittive sono state calcolate su un campione di 27 soggetti su cui sono state rilevate le variabili X = numero medio di ore dedicate all'attività fisica alla settimana e Y = indice di massa corporea

	Media	Varianza
Ore di attività fisica(X)	2.5	1.44
Indice di massa corporea(Y)	24.4	6.25

La covarianza tra numero medio di ore dedicate all'attività fisica alla settimana e indice di massa corporea è $s_{xy} = -1.8$

- (a) Determinare le stime dei minimi quadrati dei coefficienti della retta di regressione che pone l'indice di massa corporea (Y) in funzione del numero medio di ore dedicate all'attività fisica alla settimana (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 10%; (ii) calcolare il valore osservato della statistica test e prendere una decisione
- (d) Stimare l'indice di massa corporea medio per soggetti che praticano in media 4 ore di attività fisica alla settimana e costruire il relativo intervallo di confidenza al livello di confidenza del 90%.

Esercizio 3

Si consideri uno studio in cui su 27 soggetti si osservano le variabili Y = indice di massa corporea; X = numero medio di ore dedicate all'attività fisica alla settimana, e A una variabile binaria uguale a 1 per soggetti che seguono una dieta equilibrata e uguale a 0 altrimenti. Si considerino le seguenti statistiche descrittive:

Statistica	Gruppo	
	Dieta: Si	Dieta: No
n_ℓ	12	15
\bar{x}_ℓ	2.8	2.26
\bar{y}_ℓ	22.4	26
$\sum_{i=1}^{n_\ell} (x_{i\ell} - \bar{x}_\ell)^2$	8.91	8.96
$\sum_{i=1}^{n_\ell} (y_{i\ell} - \bar{y}_\ell) \cdot (x_{i\ell} - \bar{x}_\ell)$	-20.22	-13.62

- Stimare i coefficienti del modello di regressione che pone l'indice di massa corporea in funzione del numero medio di ore dedicate all'attività fisica alla settimana e della variabile binaria A (modello di analisi della covarianza senza interazione)
- Interpretare i coefficienti di regressione stimati al punto (a)
- Sapendo che la somma dei quadrati totali di Y è $SST = 105$ e che la somma dei quadrati dei residui del modello stimato al punto (a) è $SSE_e = 80$, confrontare tale modello con il modello nullo: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Scrivere la regione critica al livello di significatività del 10% (iii) Calcolare la statistica test; (iv) Prendere una decisione giustificando la risposta

Domande di Teoria

Domanda 1

Su un campione di uomini adulti si sono rilevate le seguenti variabili: Y = indice di massa corporea; X_1 = Regolare pratica dell'attività fisica (0 = No; 1 = Si); X_2 = Fascia di età (1 = Tra 18 e 40 anni; 2 = Tra 40 e 65 anni; 3 = Oltre 65 anni). La tabella sottostante mostra i risultati dell'adattamento di un modello di regressione in cui l'indice di massa corporea (Y) è posto in funzione dell'attività fisica e della fascia di età dell'umidità (relativa):

Variabile	Coefficiente
Costante	27.2
Pratica dell'attività fisica (X_1)	-1.7
Fascia di età (X_2)	
18 - 40 anni	-1.2
40 - 65 anni	-0.7

- Scrivere l'equazione di regressione stimata
- Interpretare il coefficiente relativo alla pratica dell'attività fisica
- Come dovrebbe essere interpretato il coefficiente relativo alla pratica dell'attività fisica se nel modello fosse presente l'interazione tra pratica dell'attività fisica e fascia di età?

Domanda 2

Su un campione di soggetti si osservano la variabile binaria $Y = 1$ per soggetti con indice di massa corporea superiore a 30 e $Y = 0$ per soggetti con indice di massa corporea minore o uguale a 30 e X il numero di ore dedicate in media alla settimana all'attività fisica. Su tale campione di soggetti viene stimato un modello di regressione logistica, ottenendo: $\widehat{\text{logit}}(\pi_i) = 0.2 - 0.1 \cdot x_i$. Le seguenti affermazioni contengono un errore. Individuare l'errore e spiegare perchè si tratta di un errore scrivendo l'affermazione corretta.

- Il modello di regressione logistica garantisce che il logit della probabilità assuma solo valori compresi tra zero e uno.
- La probabilità stimata che soggetti che non praticano attività fisica ($x_i = 0$) abbiano un indice di massa corporea superiore a 30 è uguale a 0.2
- In base alla stima del modello si ha che per ogni ora in più dedicata in media all'attività fisica alla settimana la probabilità che un soggetto abbia un indice di massa corporea superiore a 30 diminuisce di 0.10 punti.