

Scuola di Scienze Politiche
Corso di Laurea LM–88 SOCIOLOGIA E RICERCA SOCIALE
Compito di Metodi Statistici per la Ricerca Sociale
13 Giugno 2016

Parte applicativa

Esercizio 1

I seguenti dati si riferiscono alla percentuale di rifiuti non differenziati prodotti in un anno in un campione di comuni

Area geografica	Dimensione campionaria	Media campionaria	Varianza campionaria
Nord	20	25%	9
Centro	15	15%	12.25
Sud-Isole	15	35%	16

- (a) Calcolare la media marginale della percentuale di rifiuti non differenziati. Calcolare quindi la devianza tra gruppi e la devianza entro gruppi.
- (b) Verificare che non ci sia differenza nella percentuale di rifiuti non differenziati tra i comuni delle varie zone geografiche: (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
- (c) Il base al risultato al al punto (c), p -value per tale test sarebbe stato maggiore, minore o uguale a 0.01? Giustificare la risposta.

Esercizio 2

Sembrebbe che bere quantità moderate di vino possa prevenire l'infarto. La seguente tabella mostra alcune statistiche descrittive relative al consumo di vino (variabile X) e il numero di morti annui (morti per 100 000 persone) per malattie cardiache (variabile Y) in 19 paesi sviluppati.

$$\bar{x} = 3.03 \quad \bar{y} = 191.05 \quad s_X^2 = 6.30 \quad s_Y^2 = 4678.053 \quad s_{XY} = -144.674$$

- (a) Calcolare il coefficiente di correlazione e interpretarlo
- (b) Stimare i coefficiente del modello di regressione che pone la mortalità per malattie cardiache in funzione del consumo del vino
- (c) Sottoporre a test l'ipotesi che il coefficiente du regressione relativo al consumo di vino sia significativamente diverso da zero: (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
- (d) Per l'Italia si è osservato un consumo di vino $x = 7.9$ e una mortalità per malattie cardiache $y = 107$. Calcolare la mortalità per malattie cardiache atteso utilizzando il modello di regressione e il residuo di regressione per l'Italia. Costruire quindi un intervallo di confidenza al 99% per la mortalità per malattie cardiache atteso di paesi dove il consumo di vino è $x = 7.9$.

Esercizio 3

In uno studio sulla pratica dell'auto-palpazione come mezzo di prevenzione del cancro al seno viene stimato un modello logit in cui la pratica dell'auto-palpazione (0 = No, 1 = Si) è la variabile risposta (Y) e la

partecipazione a un corso organizzato da personale medico su come effettuare l'auto-palpazione ($X_1 = 0 = \text{No}$, $X_1 = 1 = \text{Si}$) è la variabile esplicativa. La seguente tabella mostra i risultati della stima del modello:

Variabile	Coefficiente	Errore standard
Costante	1.35	0.37
Partecipazione al corso	1.50	0.25

- Interpretare il coefficiente della variabile partecipazione al corso in termini di odds
- Costruire un intervallo di confidenza al livello di confidenza del 95% per il rapporto degli odds a favore della pratica dell'auto-palpazione. Interpretare il risultato ottenuto
- Calcolare la probabilità di successo stimata per una donna che partecipa al corso

Domande di Teoria

Domanda 1

In un recente studio viene effettuata un'analisi di regressione che pone la percentuale della vita di un bambino trascorsa in povertà in funzione del livello di istruzione della madre e la percentuale della vita del bambino trascorsa con un singolo genitore. La seguente tabella mostra i risultati dell'analisi:

Variabile	Stima	Errore standard
Costante	56.4	2.12
% vita con un solo genitore	-0.32	0.014
Livello di istruzione madre (Obbligo)		
Diploma	-1.32	0.15
Laurea	-2.53	0.27

- Scrivere l'equazione di regressione stimata
- Interpretare i coefficienti di regressione
- Scrivere l'ipotesi nulla e l'ipotesi alternativa per il confronto di tale modello con un modello (esteso) in cui è inclusa l'interazione tra le due variabili esplicative

Domanda 2

Si supponga avere dati su un campione di giovani adulti e di essere interessati a determinare le variabili rilevanti nella scelta del tempo libero. si definisca la variabile risposta Y_i tale che

$$Y_i = \begin{cases} 1 & \text{Attività fisiche} \\ 2 & \text{Attività culturali} \\ 3 & \text{Attività ricreative} \end{cases}$$

Specificare un modello di regressione logistica multinomiale considerando come categoria di riferimento le attività fisiche e come variabili esplicative il genere ($0 = \text{Maschio}$, $1 = \text{Femmina}$), l'età, il titolo di studio con categorie $1 = \text{nessun titolo o scuola dell'obbligo (riferimento)}$, $2 = \text{diploma}$, $3 = \text{laurea o più}$ e il reddito familiare.