

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Compito di Metodi Statistici per la Ricerca Sociale 09 Settembre 2015

Studente: _____

Matricola: _____

Esercizi

1. In uno studio sul consumo di droghe leggere vengono rilevate le seguenti variabili: Uso di marijuana (0 = No, 1 = Si) e uso di sigarette (0 = No, 1 = Si). La seguente tabella mostra i dati osservati:

Uso di sigarette	Uso di marijuana	
	No	Si
No	280	120
Si	390	210

- (a) Stimare le probabilità di usare marijuana fra coloro che usano sigarette e coloro che non usano sigarette.
- (b) Verificare l'ipotesi che la proporzione di coloro che usano marijuana sia maggiore tra soggetti che fumano sigarette al livello di significatività del 3%. Scrivere l'ipotesi nulla e l'ipotesi alternativa, scrivere la regione critica (di rifiuto), calcolare il valore osservato della statistica test e prendere una decisione.
- (c) Calcolare l'odds a favore dell'uso di marijuana per soggetti che usano sigarette e per soggetti che non usano sigarette. Stimare il rapporto degli odds.
- (d) Costruire un intervallo di confidenza al livello del 97% per il rapporto degli odds.
2. Un campione di 5 soggetti obesi viene sottoposto a una dieta di durata variabile. Per ciascun soggetto si osserva la durata della dieta, espressa in numero di settimane (variabile X) e il numero di chili persi durante la dieta (variabile Y). La seguente tabella riporta i valori osservati:

Variabile	Soggetto				
	1	2	3	4	5
Durata della dieta (X)	2	1	3	6	4
Numero di chili persi (Y)	5	6	7	9	8

- (a) Determinare le stime dei minimi quadrati dei coefficienti della retta di regressione che pone il numero dei chili persi (Y) in funzione della durata della dieta (X)
- (b) Sottoporre a test l'ipotesi che il coefficiente di regressione della variabile X sia significativo al livello di significatività del 5%.
- (c) Stimare il numero medio di chili che ci si attende possa essere perso da un soggetto che effettua la dieta per 5 settimane e costruire il relativo intervallo di confidenza al livello di confidenza del 95%.

3. In uno studio sull'ipertensione sono rilevate le seguenti variabili su un campione di $n = 75$ osservazioni: pressione sanguigna (variabile risposta) è posta in funzione delle seguenti variabili esplicative: Body Mass Index (BMI), età (variabile continua), genere (1 = maschi; 0 = Femmina), trattamento per l'ipertensione (1 = sì; 0 = no). La somma dei quadrati totale è $SQT = 5677.636$. Si supponga di adattare un modello di regressione che pone la pressione sanguigna in funzione delle variabili esplicative, Body Mass Index (BMI), età, genere, trattamento per l'ipertensione. Si consideri tale modello come il modello esteso. La somma dei quadrati dei residui per il modello esteso è $SQE_e = 1854.525$
- Costruire la tavola di analisi della varianza e confrontare il modello stimato con il modello nullo. Scrivere l'ipotesi nulla e l'ipotesi alternativa, scrivere la regione critica (di rifiuto), calcolare il valore osservato della statistica test e prendere una decisione.
 - Si supponga di adattare un modello ridotto in cui si pone la pressione sanguigna in funzione delle variabili esplicative, Body Mass Index (BMI) e trattamento per l'ipertensione. La somma dei quadrati dei residui per tale modello è $SQE_r = 4572.718$. Confrontare il modello ridotto con il modello esteso. Scrivere l'ipotesi nulla e l'ipotesi alternativa, scrivere la regione critica (di rifiuto), calcolare il valore osservato della statistica test e prendere una decisione.

Domande Teoriche

1. In uno studio sull'ipertensione viene stimato un modello di regressione multipla in cui la pressione sanguigna (variabile risposta) è posta in funzione delle seguenti variabili esplicative: Body Mass Index (BMI), età (variabile continua), genere (1 = maschi; 0 = Femmina), trattamento per l'ipertensione (1 = sì; 0 = no)

Variabile	Stima	Errore standard
Costante	68.15	2.59
BMI	0.58	0.06
Età	0.65	0.03
Genere: Maschio	0.94	0.59
Trattamento per l'ipertensione: Sì	6.44	0.66

- Scrivere l'equazione di regressione stimata
 - Interpretare i coefficienti di regressione
 - L'indice di determinazione lineare per tale modello è $R^2 = 0.6547$. Interpretare tale valore.
2. In uno studio sul comportamento degli individui rispetto al voto, viene stimato un modello logit che in cui si considera come variabile risposta Y la decisione di un soggetto di votare nelle prossime elezioni politiche con $Y = 1$ se l'individuo decide di votare e $Y = 0$ altrimenti e come variabile esplicativa il titolo di studio (1 = Scuola dell'obbligo o meno; 2 = Diploma di maturità; 3 = Laurea o più). La seguente tabella riporta i coefficienti stimati

Variabile	Stima	Errore standard
Costante	1.15	0.59
Titolo di studio (Obbligo)		
Diploma	0.35	0.03
Laurea o più	0.64	0.09

- Scrivere l'espressione delle probabilità stimate
- Calcolare la probabilità che un soggetto con laurea o più decida di votare
- Interpretare il coefficiente relativo al titolo di studio "Laurea o più" in termini di odds.