

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Compito di Metodi Statistici per la Ricerca Sociale 15 Dicembre 2015

Studente: _____

Matricola: _____

Esercizi

1. A un campione di 203 studenti di una certa università vengono effettuate varie domande finalizzate a valutare la qualità del servizio mensa. Sulla base delle risposte ottenute, viene calcolato un punteggio su scala da 0 a 100 che sintetizza il giudizio di ciascun intervistato. La seguente tabella mostra alcune statistiche descrittive del punteggio in gruppi di studenti classificati rispetto alla frequenza con cui usufruiscono del servizio mensa.

Uso del servizio mensa	Statistiche descrittive		
	Numero di osservazioni	Media	Varianza
1 volta a settimana	32	70	22
2-3 volte a settimana	118	75	25
4-5 volte a settimana	53	65	25

- (a) Calcolare la devianza tra gruppi e la devianza entro gruppi
- (b) Verificare l'ipotesi che non ci sia differenza tra il punteggio medio sulla qualità del servizio mensa nelle tre popolazioni definite dalle frequenze con cui è usato il servizio mensa al livello di significatività del 1%: (i) Scrivere le ipotesi nulla e alternativa, (ii) specificare la regione critica; (ii) calcolare il valore della statistica test e prendere una decisione.
- (c) Scrivere il modello di analisi della varianza che pone il punteggio sulla qualità del servizio mensa in funzione della frequenza con cui il servizio mensa è usato usare "1 volta a settimana" come livello di riferimento. Stimare quindi i parametri del modello e interpretarli.
2. I seguenti dati sono relativi a un campione di 52 studenti di scuole medie inferiori su cui si è rilevato il punteggio al test INVALSI di matematica all'esame (variabile Y) e il voto medio a matematica con cui sono ammessi all'esame di licenza media (variabile X). Entrambe le variabili sono riportate su scala 0-100.

Variabile	Media	Varianza
Voto medio matematica (X)	70	82.81
Punteggio test INVALSI di matematica (Y)	65	49

Coefficiente di correlazione: $r_{X,Y} = 0.65$

- (a) Determinare le stime dei minimi quadrati dei coefficienti della retta di regressione che pone il punteggio al test INVALSI di matematica (Y) in funzione del voto medio a matematica (X)
- (b) Stimare la varianza degli errori

- (c) Stimare il punteggio medio al test INVALSI di matematica che ci si attende possa essere ottenuto in media da studenti che hanno un voto medio a matematica di 80. Costruire quindi il relativo intervallo di confidenza al livello di confidenza del 95%.
3. In uno studio sui comportamenti di giovani adulti su un campione di $n = 30$ soggetti di età compresa tra i 20 e i 40 anni si sono rilevate le seguenti variabili Y = peso (in chilogrammi); X_1 = numero di ore settimanali dedicate allo sport o a altri esercizi fisici; X_2 = genere (0 = Femmina; 1 = Maschio); X_3 = Tipo di dieta (1 = Dieta che include carne; 2 = Vegetariana; 3 = Vegana). Si supponga di adattare un modello di regressione che pone il peso, Y in funzione delle variabili esplicative X_1, X_2, X_3 e dell'interazione tra genere e tipo di dieta ($X_2 \cdot X_3$). Si consideri tale modello come modello esteso. Per tale modello la somma dei quadrati dei residui è $SSE_e = 1693.5$. Al fine di valutare la significatività dell'interazione viene stimato un modello ridotto senza interazione per cui la somma dei quadrati dei residui è $SSE_r = 1955.7$.
- (a) Scrivere l'ipotesi nulla e l'ipotesi alternativa relativa al confronto dei due modelli.
- (b) Scrivere la regione critica al livello di significatività $\alpha = 0.05$ e la regione critica al livello di significatività $\alpha = 0.01$.
- (c) Calcolare il valore osservato della statistica test e prendere una decisione.
- (d) Il p -value per tale test è $p = 0.191$. Interpretare tale valore.

Domande Teoriche

1. In uno studio sui comportamenti di giovani adulti su un campione di $n = 120$ soggetti di età compresa tra i 20 e i 40 anni viene stimato un modello di regressione che pone il Y = peso (in chilogrammi) in funzione delle seguenti variabili X_1 = numero di ore settimanali dedicate allo sport o a altri esercizi fisici; X_2 = genere (0 = Femmina; 1 = Maschio); X_3 = Tipo di dieta (1 = Dieta che include carne; 2 = Vegetariana; 3 = Vegana).

Variabile	Stima	Errore standard
Costante	75.3	15.2
Ore di attività fisica	-0.5	0.05
Genere (Femmina)		
Maschio	10.9	3.72
Tipo di dieta (Con carne)		
Vegetariana	-0.8	0.12
Vegana	-1.1	0.58

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare i coefficienti di regressione
2. Si consideri lo studio dell'associazione tra variabili categoriche.
- (a) Completare la seguente tabella sapendo in una data popolazione che le variabili "Status occupazionale" e "Zona di residenza" sono indipendenti

Zona di residenza	Status occupazionale			Totale
	Occupato	Disoccupato	Fuori dalle forze lavoro	
Nord				180
Centro				270
Sud				90
Totale 360	120	60	540	

- (b) Quali valori assume la statistica χ^2 . In caso di indipendenza quanto vale la statistica χ^2 ?