

METODI STATISTICI PER LA RICERCA SOCIALE

CAPITOLO 12.

CONFRONTO FRA GRUPPI: L'ANALISI DELLA VARIANZA

ESERCITAZIONE

Alessandra Mattei

Dipartimento di Statistica, Informatica, Applicazioni (DiSIA)
Università degli Studi di Firenze
mattei@disia.unifi.it

LM-88 SOCIOLOGIA E RICERCA SOCIALE

Esercizio 1

I seguenti dati si riferiscono al salario annuo lordo (in migliaia di Euro) di un campione di laureati alla prima occupazione.

Gruppo j	Numerosità del gruppo n_j	Media Campionaria $\bar{x}_j$	Varianza Campionaria s_j^2
Giurisprudenza	50	18	9
Scienze Politiche	45	20.5	9.5
Economia	35	22.5	10

- (a) Calcolare la devianza tra gruppi e la devianza entro gruppi
- (b) Valutare se esiste una differenza significativa tra il salario medio dei laureati nelle varie scuole al livello di significatività $\alpha = 0.05$: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa, (ii) specificare la regione critica; (iii) calcolare il valore della statistica test e prendere una decisione.
- (c) Scrivere il modello di analisi della varianza considerando "Economia" come gruppo di riferimento e stimare i parametri del modello

Esercizio 2

In uno studio sull'abuso di alcool nelle discoteche è stato rilevato il tasso alcolemico su un campione di 104 di soggetti (maschi) all'uscita da discoteche. I 104 soggetti sono classificati per zona di residenza: Nord, Centro, Sud, Isole. Di seguito i risultati di alcune analisi

Fonte di variabilità	Somma dei Quadrati	<i>GdL</i>	Media dei Quadrati
Tra gruppi	2.54		
Entro gruppi			
Totale	3.56		

- (a) Completare la tavola di analisi della varianza
- (b) Verificare l'ipotesi che il tasso alcolemico medio sia lo stesso nelle diverse zone di residenza al livello di significatività del 1%: (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa, (ii) specificare la regione critica; (iii) calcolare il valore della statistica test e prendere una decisione.
- (c) Definire il p -valore per tale test. Il p -valore per tale test sarebbe maggiore, minore o uguale a 0.01? (Giustificare la risposta)

Esercizio 3

Su un campione di 56 liberi professionisti autonomi che operano come avvocati, commercialisti imprenditori o medici in un certo paese si è stimato un modello di analisi della varianza che pone il numero di ore lavorate (in media) alla settimana in funzione del tipo di attività (Avvocato; Commercialista; Imprenditore, Medico) e della zona in cui operano (Nord; Centro, Sud-Isole)

- (a) Completare la tavola di analisi della varianza che segue relativa a un modello di analisi della varianza a due fattori senza interazione

Fonte di variabilità	Somma dei quadrati	Gradi di libertà	Media dei quadrati	$F - value$
Regressione				
Residua	825			–
Totale	2750			–

- (b) Specificare l'ipotesi nulla e l'ipotesi alternativa a cui si riferisce il valore della statistica F calcolato al punto (a). Scrivere quindi la regione critica al livello di significatività del 1% e prendere una decisione.
- (c) Confrontare il modello in esame con un modello ridotto in cui viene eliminata la variabile zona di residenza sapendo che la somma dei quadrati dei residui del modello ridotto è $SQE_r = 912.5$. (i) Scrivere l'ipotesi nulla e l'ipotesi alternativa; (ii) Calcolare il valore della statistica test e prendere una decisione sapendo che il p -valore per tale test è 0.0082.
- (d) Scrivere il modello di analisi della varianza in presenza di interazione tra tipo di attività e zona di attività. Determinare quindi i gradi di libertà del modello.

Esercizio 4

Su una campione di $n = 30$ donne italiane con almeno un figlio vengono osservate le seguenti variabili. Y = Età al primo figlio; X_1 = Zona di residenza (1 = Sud-isole; 2 = Centro; 3 = Nord) e X_2 = Titolo di studio (1 = Nessun titolo o scuola dell'obbligo; 2 = Diploma; 3 = Laurea o più). La seguente tabella riporta le stime dei coefficienti e degli errori standard del modello di analisi della varianza che pone l'età al primo figlio in funzione della zona di residenza e del titolo di studio (le categorie di riferimento sono tra parentesi).

Variabile	Stima	Errore standard
Costante	26.8	0.724
Zona di residenza (Sud-isole)		
Centro	1.6	0.735
Nord	2.7	0.872
Livello di istruzione (Obbligo o meno)		
Diploma	1.3	0.759
Laurea o +	2.4	0.808

- (a) Scrivere l'equazione di regressione stimata
- (b) Interpretare i coefficienti di regressione
- (c) Scrivere il modello di analisi della varianza in presenza di interazione tra zona di residenza e livello di istruzione.