

**NB: Commentare sempre i risultati ottenuti**

- Si illustri il modello lineare utilizzato nell'analisi della varianza a due criteri di classificazione, precisando il significato e i vincoli dei parametri in esso contenuti.
- Un'azienda che produce stampanti laser è interessata a verificare se quattro differenti meccanismi di trascinamento dei fogli (A, B, C e D) influenzano la velocità media di stampa misurata in pagine per minuto (ppm). Per ogni tipologia di meccanismo viene estratto un campione casuale di stampanti e per ciascuna viene valutato il numero di pagine stampate in un minuto; i dati sono riportati nella seguente tabella:

A	15	16	18	16	17	16	
B	17	16	19	18	19		
C	15	15	14	13	16	15	14
D	15	16	17	18	15	16	

Sapendo che la devianza nei gruppi è 24,68 dopo aver specificato le ipotesi necessarie, utilizzando un livello di significatività  $\alpha = 0,05$ :

- si verifichi se il numero medio di stampe (in ppm)  $\mu_j$  ( $j=1, \dots, 4$ ) è uguale per i quattro differenti tipi di meccanismi;
- valutando opportunamente i risultati del punto precedente, si verifichi, ad un livello di significatività  $\alpha = 0,05$ , le seguente ipotesi:

$$H_0 : (4\mu_1 + 6\mu_4)/10 = (\mu_2 + \mu_3)/2 \quad \text{contro} \quad H_1 : (4\mu_1 + 6\mu_4)/10 \neq (\mu_2 + \mu_3)/2$$

- costruire l'intervallo di confidenza al 95% per l'effetto specifico:  $\alpha_1 = \mu_1 - \mu$ .

- Si illustrino gli obiettivi dell'analisi dei gruppi e si presentino aspetti positivi e negativi dei metodi gerarchici e non gerarchici.
- La società assicuratrice AAA, operante nel settore danni, è presente nella provincia di Milano con sei agenzie. Al fine di determinare l'ammontare dei premi da distribuire alle sei agenzie per l'attività svolta nel 2003, la società AAA è interessata a classificare le sei agenzie in gruppi sulla base di 4 parametri:  $X_1 =$  numero di polizze danni stipulate;  $X_2 =$  numero di polizze RC auto stipulate;  $X_3 =$  numero di clienti;  $X_4 =$  numero di agenti assicurativi. Di seguito è riportata la matrice dei dati standardizzati:

Agenzia	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>	Z <sub>4</sub>
A	-1,407	-1,191	-0,900	-1,364
B	1,182	0,505	1,421	1,240
C	-0,084	1,232	-0,267	0,124
D	-0,928	-1,434	-1,048	-0,992
E	1,322	0,868	1,315	1,240
F	-0,084	0,020	-0,520	-0,248

Partendo dalla matrice delle distanze euclidee tra le 6 unità ed applicando il metodo del legame medio, si è ottenuta la seguente successione di partizioni:

Partizione	Livello di distanza
{A, B, C, D, E, F}	$0 \leq d < 0,4037$
{(B, E), A, C, D, F}	$0,4037 \leq d < 0,6691$
{(B, E), (A, D), C, F}	$0,6691 \leq d < 1,2924$
{(B, E), (A, D), (C, F)}	$1,2924 \leq d < 2,592$
{(A, C, D, F), (B, E)}	$2,592 \leq d < 3,6322$
{(A, B, C, D, E, F)}	$d \geq 3,6322$

- Tracciare il dendrogramma riferito alle sei unità;
  - suggerire una opportuna partizione giustificando la scelta;
  - descrivere la partizione individuata al punto precedente.
- Si sono rilevate quattro variabili quantitative sulle 150 unità di una popolazione. Dalla matrice **R** di correlazione tra le quattro variabili, si sono ricavate la matrice diagonale **Λ** degli autovalori e la matrice **A** dei corrispondenti autovettori normalizzati.

$$\Lambda = \begin{bmatrix} 1,6652 & 0 & 0 & 0 \\ 0 & 1,4981 & 0 & 0 \\ 0 & 0 & 0,5922 & 0 \\ 0 & 0 & 0 & 0,2446 \end{bmatrix} \quad A = \begin{bmatrix} 0,643 & 0,260 & 0,492 & 0,527 \\ -0,239 & 0,690 & 0,474 & -0,492 \\ 0,715 & -0,129 & -0,148 & -0,671 \\ 0,136 & 0,663 & -0,715 & 0,174 \end{bmatrix}$$

- si determini, mediante opportuni criteri, il numero di componenti principali da mantenere nell'analisi e si valuti, con riferimento alle sole comunalità, l'idoneità della scelta effettuata;
- si calcolino e si rappresentino graficamente le correlazioni tra le variabili originarie e le componenti principali mantenute commentando opportunamente.