

**NB: Commentare sempre i risultati ottenuti**

- Analisi della varianza ad un criterio di classificazione: si illustrino le caratteristiche dello stimatore della varianza  $\sigma^2$  comune a tutte le  $k$  popolazioni basato sulla devianza nei gruppi.
- Il laboratorio di ricerca di un'azienda che produce telefoni cellulari vuole verificare se la durata media in conversazione (espressa in ore) della batteria di un telefono cellulare dipende dal tipo di software installato sul telefono ( $A, B, C$  e  $D$ ) e/o dal tipo di display del cellulare ( $I, II$  e  $III$ ).  
 A tale scopo si considerano complessivamente 48 cellulari (4 per ogni associazione tra software e display) e si valuta la durata della batteria.

La seguente tabella riporta alcune informazioni della tabella di ANOVA ottenuta con l'esperimento.

Fonte di variabilità	S.S. (Devianza)	Gradi Di Libertà			
Software	15100,25				
Display	1491,50				
Interazione	710,50				
Errore	2115,00				
<i>Totale</i>					

Dopo aver descritto il modello lineare utilizzato nell'analisi della varianza a due criteri di classificazione ed il significato dei parametri in esso contenuti, si completi la tabella ANOVA e si verifichino (commentando opportunamente i risultati ottenuti), utilizzando un livello di significatività  $\alpha = 0,025$ , le seguenti ipotesi:

- il tipo di software installato sul telefono cellulare influenza significativamente la durata media della batteria;
  - il tipo di display del telefono cellulare influenza significativamente la durata media della batteria;
  - vi è *interazione* tra software installato e display del telefono cellulare.
- La seguente tabella riporta, per cinque marche di acque oligominerali, le seguenti informazioni:  $X_1 = \text{residuo fisso (mg/l)}$ ;  $X_2 = \text{sodio (mg/l)}$ ;  $X_3 = \text{nitrato (mg/l)}$ :

Acqua	$X_1$	$X_2$	$X_3$
A	170,40	1,23	1,00
B	177,07	4,66	1,41
C	274,80	6,90	8,20
D	309,00	5,00	3,80
E	380,00	3,00	3,00

La matrice  $D$  delle *distanze euclidee* tra i tipi acqua (calcolata sui dati standardizzati), considerati nello stesso ordine di presentazione della tabella è:

$$D = \begin{bmatrix} 0 & 1,797 & 4,277 & 2,839 & 2,888 \\ 1,797 & 0 & 3,136 & 1,904 & 2,754 \\ 4,277 & 3,136 & 0 & 2,023 & 3,156 \\ 2,839 & 1,904 & 2,023 & 0 & 1,405 \\ 2,888 & 2,754 & 3,156 & 1,405 & 0 \end{bmatrix}.$$

- Tracciare il dendrogramma riferito ai tipi di acqua avvalendosi del *metodo del legame medio*;
  - suggerire una opportuna partizione, giustificando la scelta;
  - descrivere la partizione individuata al punto precedente.
- Nell'ambito di un'indagine sulle infrastrutture da potersi utilizzare per l'organizzazione di convegni e conferenze presenti sul territorio, l'ufficio statistico della provincia di Milano estrae un campione di 150 sale conferenze e rileva, per ciascuna di esse, le seguenti variabili:  $X_1 = \text{canone giornaliero per la locazione della sala (in euro)}$ ;  $X_2 = \text{distanza dall'aeroporto di Milano Malpensa (in Km)}$ ;  $X_3 = \text{distanza dalla stazione centrale dei treni (in Km)}$ ;  $X_4 = \text{capienza della sala (numero di persone)}$ . Di seguito è riportata la matrice  $C$  di correlazione tra le variabili originarie (standardizzate) e le componenti principali:

$$C = \begin{bmatrix} 0,0310 & 0,8428 & 0,5312 & 0,0805 \\ -0,8649 & 0,2066 & -0,1998 & 0,4115 \\ -0,8943 & -0,0119 & 0,1200 & -0,4309 \\ 0,1694 & 0,8379 & -0,4834 & -0,1885 \end{bmatrix}.$$

- Si stabilisca, mediante opportuni criteri, il numero di componenti principali da mantenere nell'analisi;
- si interpretino le componenti principali selezionate al punto precedente;
- si ricavi il punteggio sulle componenti principali che si è deciso di mantenere per due sale conferenze  $\mathbf{u}_1$  e  $\mathbf{u}_2$  che presentano i seguenti valori standardizzati delle 4 variabili:

unità	$Z_1$	$Z_2$	$Z_3$	$Z_4$
$\mathbf{u}_1$	0,823	-0,512	-0,309	0,574
$\mathbf{u}_2$	0,751	0,318	0,655	0,328

commentando opportunamente i risultati ottenuti