

- Si illustri il modello di analisi della varianza a due criteri di classificazione, specificando anche il contesto di applicazione e gli obiettivi.
- Una casa automobilistica impiega alcuni rappresentanti di vendita per contattare telefonicamente i concessionari. Al fine di confrontare l'efficacia di quattro differenti piani di telefonate ( $A$ ,  $B$ ,  $C$  e  $D$ ), l'ufficio vendite seleziona casualmente, per ciascun piano di telefonate,  $n_j$  ( $j = 1, \dots, 4$ ) rappresentanti. I rappresentanti seguono i rispettivi piani per 3 mesi e al termine viene rilevato, per ciascuno di essi, il fatturato espresso in migliaia di euro; i dati sono riportati nella seguente tabella:

$A$	17,25	18,4	19,55	20,7	17,25	18,4	
$B$	19,55	18,4	21,85	20,7	21,85		
$C$	17,25	17,25	16,1	14,95	18,4	17,25	16,1
$D$	20,7	18,4	17,25	18,4	19,55	18,4	

Sapendo che la *devianza totale* è 73,84 dopo aver specificato il modello da utilizzare:

- si verifichi (utilizzando un livello di significatività  $\alpha = 0,05$ ) se il fatturato medio  $\mu_j$  ( $j=1, \dots, 4$ ) è uguale per i quattro differenti piani di telefonate;
- valutando opportunamente i risultati del punto precedente, si verifichi, ad un livello di significatività  $\alpha = 0,05$ , la seguente ipotesi:

$$H_0 : \frac{\mu_1 + \mu_4}{2} = \frac{\mu_2 + \mu_3}{2} \quad \text{contro} \quad H_1 : \frac{\mu_1 + \mu_4}{2} \neq \frac{\mu_2 + \mu_3}{2}$$

interpretando opportunamente il risultato ottenuto;

- costruire l'intervallo di confidenza al 95% per la differenza:  $(\alpha_1 - \alpha_2)$ .

- Confronto tra metodi gerarchici e non gerarchici nell'analisi dei gruppi.
- L'amministrazione di un comune deve programmare degli interventi di ristrutturazione degli stabilimenti balneari. A tale scopo si ritiene utile classificare i 6 stabilimenti balneari presenti sul territorio comunale in relazione a quattro variabili:  $X_1 =$  *dimensione dello stabilimento (misurata in numero di bagnanti ospitabili)*;  $X_2 =$  *accessibilità allo stabilimento da parte di persone disabili (misurata su una scala da 0 a 10)*;  $X_3 =$  *spese di manutenzione nei tre anni precedenti (in migliaia di euro)*;  $X_4 =$  *numero di anni trascorsi dall'ultimo intervento di ristrutturazione*.

Di seguito è riportata la matrice dei dati e la matrice  $D$  delle *distanze euclidee* tra i sei stabilimenti, calcolata sui dati standardizzati, considerati nello stesso ordine di presentazione della tabella:

Stabilimento	$X_1$	$X_2$	$X_3$	$X_4$
$A$	115	6	15	2
$B$	85	8	25	1
$C$	150	7	20	2
$D$	60	5	3	3
$E$	200	9	30	1
$F$	120	4	4	5

$$D = \begin{bmatrix} 0 & 1,826 & 1,094 & 1,947 & 3,068 & 2,710 \\ & 0 & 1,791 & 3,209 & 2,670 & 4,350 \\ & & 0 & 2,960 & 2,031 & 3,290 \\ & & & 0 & 4,950 & 2,062 \\ & & & & 0 & 5,186 \\ & & & & & 0 \end{bmatrix}$$

- Tracciare il dendrogramma riferito ai sei stabilimenti avvalendosi del *metodo del legame completo*;
  - suggerire una opportuna partizione, giustificando la scelta;
  - descrivere la partizione individuata al punto precedente.
- Si presentino i criteri per stabilire il numero di componenti principali da mantenere nell'analisi e gli strumenti utili alla verifica della scelta effettuata.