

NB: Commentare sempre i risultati ottenuti

1. Un'azienda che produce scope elettriche sta progettando un nuovo modello e vuole verificare se due differenti dispositivi A e B influenzano significativamente la capacità di aspirazione (misurata in litri per secondo) su differenti superfici (*piastrelle, moquette e parquet*). A tale proposito, per ogni associazione tra dispositivo e superficie, vengono testate 4 scope elettriche.

La seguente tabella riporta le capacità di aspirazione medie in litri per secondo (Lt/sec):

Superficie	Dispositivo		$\bar{X}_{j..}$
	A	B	
<i>Piastrelle</i>	42,00	40,50	41,25
<i>Moquette</i>	37,75	39,75	38,75
<i>Parquet</i>	41,00	39,00	40,00
$\bar{X}_{.k.}$	40,25	39,75	40,00

Sapendo che la *devianza totale* è 60,00 dopo aver specificato le ipotesi necessarie, si verifichi, ad un livello di significatività $\alpha = 0,05$, se:

- a) la superficie influenza significativamente la capacità di aspirazione media;
 - b) il dispositivo influenza significativamente la capacità di aspirazione media;
 - c) vi è *interazione* tra superficie e dispositivo.
2. La seguente tabella riporta i dati relativi alle infrastrutture di alcuni aeroporti italiani (fonte ISTAT - Settore Servizi Statistiche del trasporto aereo Anno 2001): $X_1 =$ estensione (in ettari); $X_2 =$ distanza dalla città (in km); $X_3 =$ area parcheggio aerei (in mq); $X_4 =$ numero di piste.

Aeroporto	X_1	X_2	X_3	X_4
BG Orio al Serio	300	5	176000	2
MI Linate	385	8	304000	2
MI Malpensa	1244	48	1396000	2
ROMA Fiumicino	1605	34	797250	4
TO Caselle	288	15	175000	1
VE Tessera	330	12	123000	2

La matrice D delle *distanze euclidee* tra i sei aeroporti, calcolata sui dati standardizzati, considerati nello stesso ordine di presentazione della tabella è:

$$D = \begin{bmatrix} 0 & 0,374 & 4,227 & 4,045 & 1,288 & 0,470 \\ & 0 & 3,859 & 3,774 & 1,248 & 0,481 \\ & & 0 & 2,815 & 4,003 & 3,997 \\ & & & 0 & 4,548 & 3,862 \\ & & & & 0 & 1,139 \\ & & & & & 0 \end{bmatrix}$$

- a) Tracciare il dendrogramma riferito ai sei aeroporti avvalendosi del *metodo del legame completo*;
 - b) suggerire una opportuna partizione, giustificando la scelta;
 - c) descrivere la partizione individuata al punto precedente.
3. Sulle 100 unità di una popolazione si sono rilevate cinque variabili quantitative. Dalla matrice R di correlazione tra le cinque variabili, si sono estratte le componenti principali. Di seguito è riportata la matrice C di correlazione tra le variabili originarie (righe) e le componenti principali (colonne):

$$C = \begin{bmatrix} -0,8105 & -0,3290 & 0,4825 & -0,0450 & 0,0003 \\ -0,8566 & -0,3810 & -0,3275 & -0,1173 & 0,0003 \\ -0,9019 & -0,3889 & -0,1097 & 0,1522 & 0,0001 \\ 0,7911 & -0,5974 & 0,0074 & 0,0001 & -0,1315 \\ 0,7938 & -0,5936 & 0,0072 & 0,0004 & 0,1319 \end{bmatrix}.$$

- a) Si determini, mediante opportuni criteri, il numero di componenti principali da mantenere nell'analisi;
 - b) si valuti, con riferimento alle sole comunalità, l'idoneità della scelta effettuata.
4. Si dimostri, limitatamente alla prima componente principale, la relazione fra le varianze delle componenti principali estraibili da una matrice Z di dati standardizzati riferita ad n unità e p variabili e gli autovalori della matrice di correlazione fra le variabili stesse.