

NB: Commentare sempre i risultati ottenuti

1. Tabelle e matrici riferite a dati statistici multidimensionali.
2. Un'azienda è interessata a verificare se la qualità dei macchinari che produce e cede in leasing dipende dallo stabilimento in cui il macchinario viene prodotto (A, B e C) e dal processo produttivo in cui il macchinario viene impiegato (I, II e III). Per ogni associazione tra stabilimento di produzione e processo produttivo di impiego vengono selezionati casualmente 5 macchinari e, per ciascuno di essi, viene rilevato il numero di interventi di manutenzione effettuati negli ultimi 3 anni. La seguente tabella riporta il numero medio di interventi di manutenzione:

Stabilimento	Processo produttivo			$\bar{X}_{j..}$
	I	II	III	
A	4,200	3,200	6,000	4,467
B	4,400	1,200	4,000	3,200
C	2,800	3,800	5,600	4,067
$\bar{X}_{.k.}$	3,800	2,733	5,200	3,911

Sapendo che la *devianza totale* è 277,644 dopo aver specificato le ipotesi necessarie, si verifichi, ad un livello di significatività $\alpha = 0,05$, se:

- a) lo stabilimento di produzione influenza significativamente il numero di interventi di manutenzione;
 - b) il processo produttivo di impiego influenza significativamente il numero di interventi di manutenzione;
 - c) vi è *interazione* tra stabilimento di produzione e processo produttivo di impiego;
 - d) qualora si sia rifiutata una delle ipotesi precedenti, costruire un opportuno contrasto lineare e verificarne l'ipotesi di nullità.
3. Dopo aver descritto gli obiettivi dell'analisi delle componenti principali, si definisca la prima componente principale e si illustri il procedimento che porta alla determinazione della stessa.
 4. Per cinque alberghi di una nota località sciistica, sono state rilevate 4 variabili: $X_1 =$ numero di posti letto; $X_2 =$ numero di servizi aggiuntivi (piscina, sauna, palestra, sala giochi, ecc...); $X_3 =$ prezzo di una notte e prima colazione (euro); $X_4 =$ distanza dalle piste da sci (km). Di seguito è riportata la matrice dei dati standardizzati:

Albergo	Z ₁	Z ₂	Z ₃	Z ₄
A	-0,5590	-0,4423	-1,4985	0,9267
B	1,1180	1,7693	-0,5620	1,3296
C	0,5590	0,2949	-0,0937	-0,0806
D	-1,6771	-1,1795	1,3112	-1,0879
E	0,5590	-0,4423	0,8429	-1,0879

La matrice delle *distanze euclidee* tra i cinque alberghi, considerati nello stesso ordine di presentazione della precedente tabella, è:

$$D = \begin{bmatrix} 0 & 2,9569 & 2,1867 & 3,7076 & 3,2850 \\ & 0 & 2,1666 & 5,0854 & 3,6086 \\ & & 0 & 3,1878 & 1,5605 \\ & & & 0 & 2,4006 \\ & & & & 0 \end{bmatrix}$$

- a) Tracciare il dendrogramma riferito alle cinque unità avvalendosi del *metodo del legame completo*;
 - b) suggerire una opportuna partizione, giustificando la scelta;
 - c) descrivere la partizione individuata al punto precedente.
5. La seguente matrice **C** riporta i coefficienti di correlazione lineare tra quattro variabili standardizzate (righe) e le quattro componenti principali da esse estratte (colonne):

$$C = \begin{bmatrix} 0,6875 & 0,6221 & 0,3679 & -0,0708 \\ -0,9581 & 0,2329 & -0,0186 & -0,1655 \\ 0,8501 & -0,4519 & -0,2343 & -0,1348 \\ -0,1973 & -0,9104 & 0,3629 & -0,0238 \end{bmatrix}$$

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.