

NB: Commentare sempre i risultati ottenuti

1. Data una matrice $X_{(n \times p)}$ individui-caratteri, riferita ad n unità statistiche sulle quali sono state rilevate p variabili quantitative, si illustri il procedimento per ricavare la matrice $Z_{(n \times p)}$ dei dati standardizzati e la matrice $R_{(p \times p)}$ di correlazione fra le p variabili.
2. Dopo aver illustrato gli obiettivi dell'analisi dei gruppi, si presentino i metodi gerarchici aggregativi del legame singolo e del legame completo.
3. E' noto che il carattere quantitativo X , presente su quattro popolazioni distinte, ha distribuzione normale di valore atteso μ_j ($j = 1, 2, 3, 4$) e varianza σ^2 identica nelle quattro popolazioni. Si è interessati al contrasto lineare:

$$L = 1,25\mu_1 - 0,7\mu_2 + 0,25\mu_3 - 0,8\mu_4.$$

A tale scopo si seleziona in ciascuna delle quattro popolazioni un campione casuale di numerosità n .

- a) Si consideri lo stimatore *naturale* per L e se ne determini il valore atteso e la varianza;
 - b) si individui la distribuzione dello stimatore individuato al punto precedente;
 - c) si costruisca la statistica test per verificare l'ipotesi nulla $H_0 : L = 0$ contro l'alternativa $H_1 : L \neq 0$ ad un livello di significatività α .
4. Per sei aziende sono state rilevate tre variabili: $X_1 = \text{numero di dipendenti}$; $X_2 = \text{fatturato dell'anno 2003}$; $X_3 = \text{numero di filiali}$. Di seguito è riportata la matrice dei dati standardizzati.

Azienda	Z_1	Z_2	Z_3
A	0,506	0,412	0,878
B	-0,744	-0,559	-0,293
C	0,952	0,764	0,293
D	-0,833	-0,823	-0,878
E	1,398	1,558	1,464
F	-1,279	-1,352	-1,464

- a) Applicare la procedura di analisi dei gruppi non gerarchica delle k -medie selezionando come centri iniziali le aziende **A** e **C**;
 - b) descrivere la partizione ottenuta.
5. Una banca, nell'ambito del lancio di un nuovo prodotto finanziario, effettua uno studio su 150 dei suoi migliori clienti. Per ciascun cliente si sono rilevate le seguenti variabili: $X_1 = \text{giacenza media annua del conto corrente}$; $X_2 = \text{spese variabili di gestione del conto corrente}$; $X_3 = \text{numero di movimenti effettuati nell'ultimo anno}$; $X_4 = \text{investimento medio annuo in fondi azionari}$. Di seguito è riportata la matrice **C** di correlazione tra le variabili originarie e le componenti principali:

$$C = \begin{bmatrix} 0,9650 & 0,0648 & -0,0224 & 0,2531 \\ 0,1226 & -0,8381 & -0,5311 & -0,0237 \\ 0,4018 & 0,7932 & -0,4381 & -0,1325 \\ 0,8723 & -0,3193 & 0,3012 & -0,2156 \end{bmatrix}$$

- a) 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;
- b) si interpretino le componenti principali selezionate al punto precedente;
- c) si rappresentino graficamente le correlazioni tra le variabili originarie e le componenti principali mantenute commentando opportunamente.