

NB: Commentare sempre i risultati ottenuti

1. Analisi della varianza a due criteri di classificazione: si illustri la scomposizione della devianza **fra** i gruppi e si presenti la costruzione della statistica test per verificare l'ipotesi nulla che tutti gli effetti specifici imputabili ai c livelli di un fattore siano nulli contro l'alternativa che almeno uno di essi sia significativamente diverso da zero.
2. Dalla tabella di composizione degli alimenti (Fonte Inran 2003) si sono ricavate le seguenti informazioni per alcuni tipi di agrumi: X_1 zuccheri (in grammi); X_2 fibra totale (in grammi); X_3 energia (in Kcal) e X_4 vitamina C (in grammi).

I dati, per 100 grammi commestibili, sono riportati nella seguente tabella:

Agrumi	X_1	X_2	X_3	X_4
Mandarini	17,6	1,7	72	42
Mandaranci	12,8	2,2	53	37
Arance	7,8	1,6	34	50
Clementine	8,7	1,2	37	54
Limoni	2,3	1,9	11	50
Pompelmi	6,2	1,6	26	40

La matrice D delle *distanze euclidee* calcolata sui dati standardizzati è:

$$D = \begin{bmatrix} 0 & 2,294 & 3,109 & 3,614 & 4,673 & 3,360 \\ 2,294 & 0 & 3,213 & 4,441 & 3,839 & 2,804 \\ 3,109 & 3,213 & 0 & 1,481 & 1,908 & 1,707 \\ 3,614 & 4,441 & 1,481 & 0 & 3,031 & 2,733 \\ 4,673 & 3,839 & 1,908 & 3,031 & 0 & 2,200 \\ 3,360 & 2,804 & 1,707 & 2,733 & 2,200 & 0 \end{bmatrix}.$$

- a) Tracciare il dendrogramma riferito ai sei agrumi avvalendosi del *metodo del legame medio*;
 - b) suggerire una opportuna partizione, giustificando la scelta;
 - c) descrivere la partizione individuata al punto precedente.
3. Dopo aver fornito la definizione generale di misura di similarità, si presenti il concetto di concordanza e di discordanza illustrando il loro impiego nella costruzione di misure di similarità ricorrendo anche ad opportuni esempi.
 4. Sulle unità di una popolazione si sono rilevate quattro variabili quantitative. Dalla matrice R di correlazione tra le quattro variabili, si sono ricavati gli autovalori λ_j ($j = 1, \dots, 4$) e la matrice A dei corrispondenti autovettori normalizzati:

$$\begin{array}{l} \lambda_1 = 1,8685 \\ \lambda_2 = 1,4376 \\ \lambda_3 = 0,5652 \\ \lambda_4 = 0,1287 \end{array} \quad A = \begin{bmatrix} -0,7059 & -0,0540 & -0,0298 & 0,7056 \\ -0,0897 & 0,6990 & -0,7064 & -0,0661 \\ -0,2939 & -0,6615 & -0,5827 & -0,3694 \\ -0,6381 & 0,2663 & 0,4006 & -0,6011 \end{bmatrix}.$$

- a) Si stabilisca, mediante opportuni criteri, il numero di componenti principali da mantenere nell'analisi;
- b) si valuti, ricorrendo al calcolo delle comunalità, l'idoneità della scelta effettuata;
- c) si rappresentino graficamente le correlazioni tra le variabili originarie e le prime due componenti principali e si riporti, nello stesso grafico, l'unità statistica il cui punteggio sulle prime due componenti principali è rispettivamente 0,82 e $-0,12$ interpretandone la posizione rispetto alle variabili originarie.