

NB: Commentare sempre i risultati ottenuti

1. Una società presente su internet con un sito di informazioni e servizi sui propri prodotti vuole valutare il gradimento da parte dei visitatori del sito stesso. A tale scopo si decide di predisporre un questionario da sottoporre all'utente appena accede al sito. Al fine di scegliere quale questionario, tra i tre disponibili (*I*, *II* e *III*), sia più opportuno impiegare, ciascun questionario è sottoposto a rotazione sul sito nel corso di una settimana.

Per ciascun questionario sottoposto si valuta se il questionario è stato compilato interamente o se l'utente ha abbandonato o non ha iniziato la compilazione del questionario.

La seguente tabella riporta per ciascun giorno, le percentuali di questionari compilati interamente:

<i>I</i>	35	36	38	42	35	40	33
<i>II</i>	23	30	34	32	31	32	33
<i>III</i>	38	44	43	50	39	41	35

Sapendo che la *devianza totale* è 686,95 dopo aver specificato le ipotesi e il modello da utilizzare:

- a) si verifichi (utilizzando un livello di significatività $\alpha = 0,05$) se la percentuale media di questionari compilati è significativamente influenzata dal tipo di questionario;
- b) indicata con μ_j ($j = 1, 2, 3$) la percentuale media di questionari compilati per il questionario j , si verifichi, ad un livello di significatività $\alpha = 0,05$, la seguente ipotesi:

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

interpretando opportunamente il risultato ottenuto;

- c) si costruisca l'intervallo di confidenza al 98% per: $\alpha_1 = \mu_1 - \mu$.

2. Per alcuni porti italiani si sono considerate le merci e i passeggeri trasportati in navigazione internazionale nell'anno 2001. La seguente tabella riporta: X_1 = sbarchi di merce in migliaia di tonnellate; X_2 = imbarchi di merce in migliaia di tonnellate; X_3 = sbarchi di passeggeri (in migliaia); X_4 = imbarchi di passeggeri (in migliaia).

Porto	X_1	X_2	X_3	X_4
Bari	1800	634	548	555
Venezia	16811	2275	271	273
Ancona	3797	1545	604	609
Savona - Vado	11238	456	166	179
Livorno	10805	2801	352	335
Brindisi	6949	893	429	430

Fonte: Istat - Annuario statistico italiano 2003

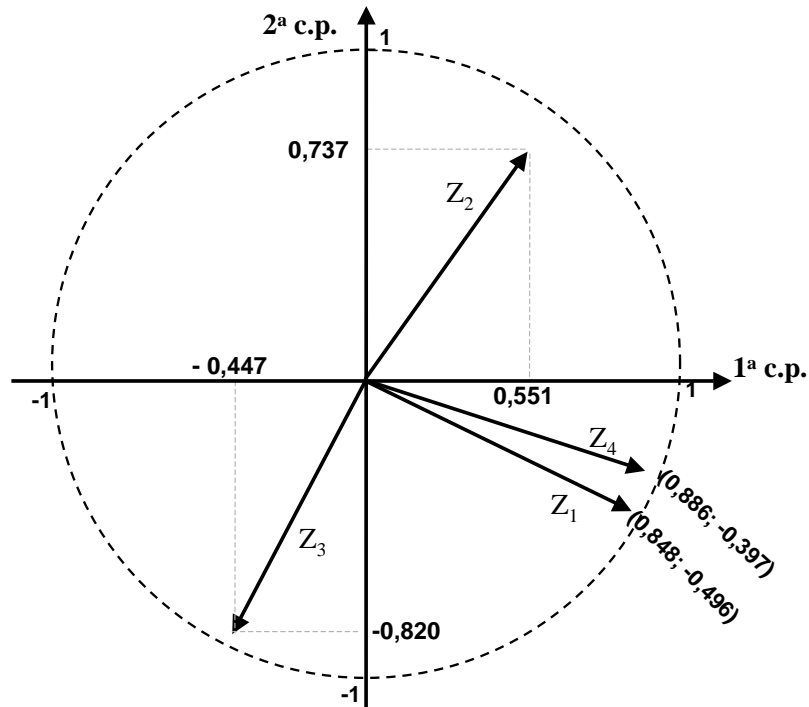
La matrice D delle *distanze euclidee* calcolata sui dati standardizzati è:

$$D = \begin{bmatrix} 0 & 4,399 & 1,238 & 4,010 & 3,645 & 1,562 \\ 4,399 & 0 & 4,144 & 2,555 & 1,502 & 2,929 \\ 1,238 & 4,144 & 0 & 4,493 & 3,177 & 1,921 \\ 4,010 & 2,555 & 4,493 & 0 & 3,153 & 2,596 \\ 3,645 & 1,502 & 3,177 & 3,153 & 0 & 2,474 \\ 1,562 & 2,929 & 1,921 & 2,596 & 2,474 & 0 \end{bmatrix}$$

- a) Tracciare il dendrogramma riferito ai porti avvalendosi del *metodo del legame singolo*;
 - b) suggerire una opportuna partizione, giustificando la scelta;
 - c) descrivere la partizione individuata al punto precedente.
3. Data una matrice X individui-caratteri, si illustri analiticamente la procedura per la determinazione della prima componente principale e se ne determini la varianza.

CONTINUA SUL RETRO

4. Sulle 100 unità di una popolazione sono state rilevate quattro variabili quantitative. Dalla matrice R di correlazione tra le quattro variabili, si sono ricavate le componenti principali. A seguito dell'applicazione dei consueti criteri di selezione, si è deciso di mantenere le prime due componenti principali. Si riporta di seguito la rappresentazione grafica delle correlazioni tra le variabili originarie standardizzate ($Z_j, j = 1, \dots, 4$) e le prime due componenti principali:



- si calcolino le comunalità per ciascuna delle quattro variabili;
- si calcolino gli autovettori associati alle prime due componenti principali;
- si commenti la rappresentazione grafica delle correlazioni tra le variabili originarie e le prime due componenti principali.