

NB: Commentare sempre i risultati ottenuti

1. L'ufficio cultura di un comune è interessato a verificare se la spesa delle famiglie in tecnologia (computer, lettori dvd, fotocamera digitale, abbonamento adsl ecc...) dipende dal titolo di studio del capofamiglia e dalla presenza di figli nel nucleo familiare. A tale scopo l'universo delle famiglie residenti nel comune al 31.12.05 viene suddiviso in base ai due caratteri: titolo di studio del capofamiglia (licenza elementare, licenza media, diploma, laurea) e presenza di figli di almeno 12 anni nel nucleo familiare (si, no). Da ciascuna delle popolazioni costituite in base alle associazioni tra i due caratteri viene estratto un campione casuale di 4 famiglie e viene rilevata la spesa (in euro) sostenuta nel 2005 per l'acquisto di apparecchiature e servizi di tecnologia. La seguente tabella riporta le spese medie:

Figli di almeno 12 anni	Tit. di studio del C.F.				$\bar{X}_{j..}$
	<i>lic. elem.</i>	<i>lic. media</i>	<i>diploma</i>	<i>laurea</i>	
S_i	1350	1425	1750	1875	1600,00
No	900	1100	1150	1312	1115,50
$\bar{X}_{.k.}$	1125,00	1262,50	1450,00	1593,50	1357,75

Sapendo che la devianza nei gruppi è pari a 1053678, dopo aver specificato il modello da utilizzare, considerando un livello di significatività $\alpha = 0,05$, si verifichi se:

- il fattore *titolo di studio del capofamiglia* influenza significativamente la spesa media per l'acquisto di apparecchiature e servizi di tecnologia;
 - il fattore *presenza di figli di almeno 12 anni* influenza significativamente la spesa media per l'acquisto di apparecchiature e servizi di tecnologia;;
 - vi è *interazione* tra i due fattori.
2. Un'azienda vende enciclopedie tramite un'attività di telemarketing. Al fine di valutare la performance di 6 operatori telefonici assunti nell'ultimo trimestre del 2005, si rileva, per ciascun operatore: $X_1 = \text{numero di enciclopedie vendute a nuovi clienti}$; $X_2 = \text{numero di aggiornamenti venduti a clienti che avevano acquistato l'enciclopedia negli anni precedenti}$; $X_3 = \text{numero di potenziali clienti che non hanno acquistato l'enciclopedia ma che hanno dimostrato interesse all'offerta lasciando i propri dati}$.

Di seguito è riportata la matrice dei dati:

Operatore	X_1	X_2	X_3
A	14	11	3
B	8	7	15
C	10	9	9
D	7	8	14
E	13	11	4
F	9	11	9

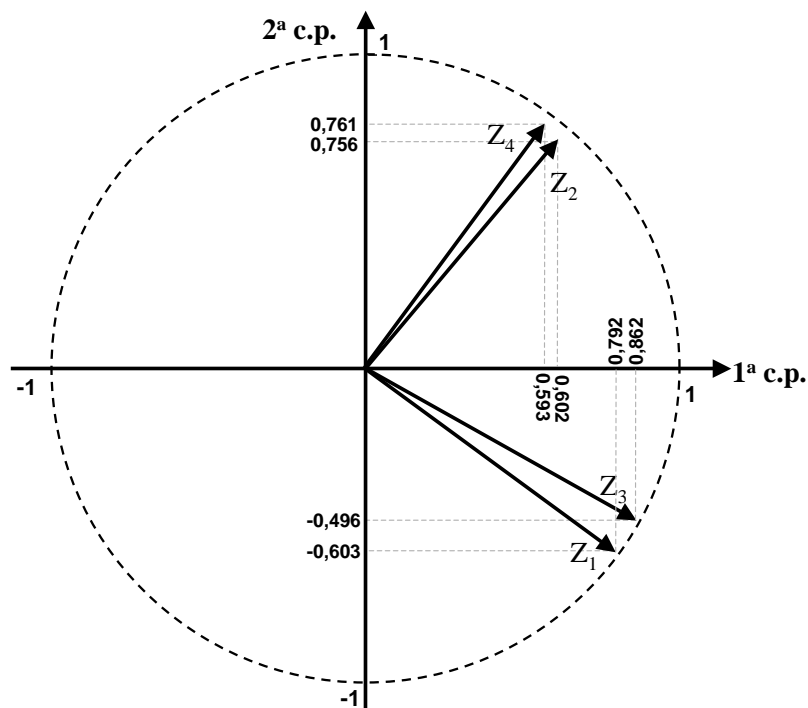
La matrice delle *distanze euclidee* tra i sei operatori calcolata sui dati originari, considerati nello stesso ordine di presentazione della precedente tabella, è:

$$D = \begin{bmatrix} 0 & 14,000 & 7,483 & 13,379 & 1,414 & 7,810 \\ & 0 & 6,633 & 1,732 & 12,728 & 7,280 \\ & & 0 & 5,916 & 6,164 & 2,236 \\ & & & 0 & 12,042 & 6,164 \\ & & & & 0 & 6,403 \\ & & & & & 0 \end{bmatrix}$$

- Tracciare il dendrogramma riferito ai sei operatori avvalendosi del *metodo del centroide*;
 - suggerire una opportuna partizione, giustificando la scelta;
 - descrivere la partizione individuata al punto precedente.
3. Data una matrice $X_{(n \times p)}$ che riporta la rilevazione di p caratteri quantitativi su n unità statistiche ($n > p$), si illustri analiticamente la procedura per la determinazione della prima componente principale.

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 calcoli la varianza di ciascuna delle componenti principali;
- si individui quale delle quattro variabili originarie è meglio riprodotta dalle prime due componenti principali;
- si commenti la rappresentazione grafica delle correlazioni tra le variabili originarie e le prime due componenti principali.