

1. Un'azienda che produce detersivi per bucato in lavatrice è interessata a testare l'efficacia di 4 nuovi detersivi (A , B , C e D). A tale scopo ciascuno dei detersivi viene testato a due differenti livelli di temperatura: *freddo* e *caldo*. Per ogni associazione tra livello di temperatura e detersivo viene considerato un campione di $n = 3$ bucati e, per ciascuno di essi, al termine del bucato, viene misurata mediante opportuni strumenti, la percentuale X di sporco eliminata.

Nella seguente tabella sono riportate le percentuali medie:

Livello di temperatura	Detersivo				$\bar{X}_{j.}$
	A	B	C	D	
<i>freddo</i>	92,4	95,8	96,2	92,8	94,3
<i>caldo</i>	96,4	97,8	94,8	97,8	96,7
$\bar{X}_{.k.}$	94,4	96,8	95,5	95,3	95,5

Sapendo che la devianza totale è pari a 127,58, dopo aver specificato il modello da utilizzare, considerando un livello di significatività $\alpha = 0,05$, si verifichi se:

- il fattore *livello di temperatura* influenza significativamente la percentuale media di sporco eliminata;
 - il fattore *detersivo* influenza significativamente la percentuale media di sporco eliminata;
 - vi è interazione tra *livello di temperatura* e *detersivo*.
2. La seguente tabella riporta alcune informazioni rilevate dall'Istat con l'indagine *Aspetti della vita quotidiana* svolta nel febbraio 2006 e relative alla disponibilità di nuove tecnologie presso le famiglie. In particolare, per alcune regioni, si sono considerate le percentuali di famiglie che possiedono: $X_1 =$ la TV digitale; $X_2 =$ l'antenna parabolica; $X_3 =$ il decoder digitale terrestre; $X_4 =$ la TV via cavo.

Regione	X_1	X_2	X_3	X_4
Marche	5,4	28,6	17,8	4,8
Lazio	5,8	28,5	19,4	4,3
Abruzzo	5,1	27,9	14,7	5,1
Veneto	4,2	23,9	13,2	2,7
Calabria	4,4	27,7	13,4	2,9
Sicilia	4,1	21,6	10,4	3,5

La matrice delle *distanze euclidee* tra le sei regioni calcolata sui dati standardizzati, considerati nello stesso ordine di presentazione della precedente tabella, è:

$$D = \begin{bmatrix} 0 & 0,988 & 1,210 & 3,778 & 3,006 & 4,377 \\ & 0 & 2,114 & 4,072 & 3,351 & 4,853 \\ & & 0 & 3,375 & 2,681 & 3,629 \\ & & & 0 & 1,480 & 1,552 \\ & & & & 0 & 2,628 \\ & & & & & 0 \end{bmatrix}$$

- Tracciare il dendrogramma riferito alle sei regioni avvalendosi del *metodo del legame singolo*;
 - suggerire una opportuna partizione, giustificando la scelta;
 - descrivere la partizione individuata al punto precedente.
3. Data una matrice X che riporta la rilevazione di p caratteri quantitativi su n unità statistiche ($n > p$), si illustri analiticamente come ricavare la matrice Z dei dati standardizzati.

CONTINUA SUL RETRO

4. Dall'annuario ISTAT del 2005 si è ricavata, per ciascuna regione, la distribuzione percentuale della popolazione residente di 15 anni e oltre classificata in base al titolo di studio: $X_1 = \text{Laurea, dottorato o diploma universitario}$; $X_2 = \text{maturità}$; $X_3 = \text{qualifica professionale}$ e $X_4 = \text{licenza media}$. 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:

$$\begin{array}{l} \lambda_1 = 2,4086 \\ \lambda_2 = 1,1184 \\ \lambda_3 = 0,3653 \\ \lambda_4 = 0,1077 \end{array} \quad A = \begin{bmatrix} 0,531 & 0,378 & 0,620 & 0,436 \\ 0,609 & -0,183 & 0,125 & -0,761 \\ -0,173 & 0,900 & -0,135 & -0,377 \\ -0,563 & -0,118 & 0,762 & -0,297 \end{bmatrix}.$$

- si stabilisca, mediante opportuni criteri, il numero di componenti principali da mantenere nell'analisi;
- si interpretino le componenti principali selezionate al punto precedente;
- si rappresentino graficamente le correlazioni tra le variabili originarie e le prime due componenti principali;
- si ricavi il punteggio sulle prime due componenti principali per le seguenti due regioni che presentano i seguenti valori standardizzati delle 4 variabili:

Regione	Z_1	Z_2	Z_3	Z_4
Trentino Alto Adige	-0,620	-1,649	2,672	1,085
Lazio	2,591	2,643	-0,301	-0,763

commentando opportunamente i risultati ottenuti.