

NB: Commentare sempre i risultati ottenuti

1. Un'azienda è specializzata nell'offerta di prodotti assicurativi attraverso telefono e internet. In merito alle polizze RC auto, le denunce dei sinistri possono essere inviate dai clienti via posta ordinaria (*I*), via fax (*II*) o via email (*III*) presso una delle due sedi operative dell'azienda (*Milano* e *Roma*). La pratica del sinistro viene correttamente aperta solo quando l'azienda ha ricevuto tutti i documenti e le informazioni necessarie. Al fine di verificare se i tempi medi di apertura delle pratiche dipendono dalla modalità di invio della documentazione e/o dalla sede operativa che li riceve, si estrae, per ogni associazione tra i due caratteri, un campione di 4 pratiche sinistri e se ne rileva il tempo di apertura in giorni lavorativi (tempo intercorrente tra la denuncia del sinistro all'assicurazione e l'apertura). La seguente tabella riporta i tempi medi di apertura delle pratiche sinistri:

Modalità invio	Sede operativa		$\bar{X}_{j..}$
	<i>Milano</i>	<i>Roma</i>	
<i>I</i>	11,00	14,25	12,625
<i>II</i>	6,00	7,50	6,75
<i>III</i>	5,00	3,00	4,00
$\bar{X}_{.k.}$	7,3	8,25	7,792

- a) Dopo aver specificato il modello da utilizzare si calcolino le stime per gli effetti specifici $\alpha_j = \mu_j - \mu_{..}$ $j = 1, 2, 3$ imputabili alle modalità del fattore *modalità di invio* commentando opportunamente;
- b) sapendo che la *devianza totale* è 403,9583, si verifichino ad un livello di significatività $\alpha = 0,05$ le seguenti ipotesi:
- b1)** la modalità di invio della documentazione influenza significativamente il tempo medio di apertura della pratica;
 - b2)** la sede operativa influenza significativamente il tempo medio di apertura della pratica;
 - b3)** vi è *interazione* tra modalità di invio e sede operativa.
2. Considerate due unità statistiche presso le quali siano stati rilevati p caratteri dicotomici, si illustri come valutarne la similarità attraverso il concetto di concordanza e discordanza.
3. Un comune sta svolgendo uno studio sulla situazione del traffico e della pericolosità delle sei principali arterie stradali del proprio territorio. Per ciascuna delle sei strade si considerano quattro variabili riferite alle rilevazioni effettuate nel mese di giugno 2005: $X_1 =$ *lunghezza del tratto stradale nel territorio comunale in Km*; $X_2 =$ *numero di corsie*; $X_3 =$ *numero medio di veicoli transitati in un giorno*; $X_4 =$ *numero medio di incidenti in un giorno*. Di seguito è riportata la matrice dei dati standardizzati:

Strada	Z_1	Z_2	Z_3	Z_4
A	1,182	0,505	1,421	1,240
B	-1,407	-1,191	-0,900	-1,364
C	-0,084	1,232	-0,267	0,124
D	1,322	0,868	1,315	1,240
E	-0,084	0,020	-0,520	-0,248
F	-0,928	-1,434	-1,048	-0,992

Partendo dalla matrice delle distanze city-block tra le 6 unità ed applicando il *metodo del centroide*, si sono ottenute le seguenti partizioni iniziali della successione delle partizioni:

Partizione	Livello di distanza
$\{A, B, C, D, E, F\}$	$0 \leq d < 0,609$
$\{(A, D), B, C, E, F\}$	$0,609 \leq d < 1,242$
$\{(B, F), (A, D), C, E\}$	$1,242 \leq d < 1,837$

- a) Si completi la successione delle partizioni e si tracci il dendrogramma riferito alle sei unità;
- b) si suggerisca una opportuna partizione giustificando la scelta;

c) si descriva la partizione individuata al punto precedente.

4. Sulle 100 unità di una popolazione si sono rilevate quattro variabili quantitative. Dalla matrice R di correlazione tra le quattro variabili, si sono ricavate le componenti principali.

Di seguito è riportata la matrice C di correlazione tra le variabili originarie e le componenti principali:

$$C = \begin{bmatrix} -0,8598 & 0,0044 & 0,5101 & -0,0236 \\ -0,1127 & -0,9840 & -0,0070 & 0,1377 \\ -0,9031 & -0,2451 & -0,2825 & -0,2111 \\ -0,8750 & 0,3754 & -0,2087 & 0,2234 \end{bmatrix}$$

a) Si determini, mediante opportuni criteri, il numero di componenti principali da mantenere nell'analisi e se ne calcolino le comunalità al fine di verificare se la scelta effettuata è idonea;

b) si "interpretino" le componenti principali selezionate al punto precedente;

c) si ricavi il punteggio sulle componenti principali, che si è deciso di mantenere, per l'unità \mathbf{u}_1 che presenta i seguenti valori standardizzati delle 4 variabili:

unità	Z_1	Z_2	Z_3	Z_4
\mathbf{u}_1	0,823	-0,512	0,309	-0,574