

- Analisi della varianza ad un criterio di classificazione: definizione ed impieghi di un contrasto lineare. Si proponga un adeguato stimatore per un contrasto lineare e se ne illustrino le caratteristiche.
- Una compagnia petrolifera che gestisce una rete di stazioni per il rifornimento di carburante sul territorio nazionale, ha attivato nel 2006 una campagna promozionale secondo due modalità differenti: tessera a *raccolta punti* per il ritiro di premi al raggiungimento di una determinata soglia oppure cartolina *gratta e vinci* consegnata al cliente che effettua un rifornimento di carburante per un importo di almeno 30 euro. I singoli distributori hanno optato per una o per l'altra modalità della campagna promozionale. Al fine di verificare l'efficacia della campagna promozionale, anche in relazione all'ubicazione dei distributori (*autostrada*, *strada extraurbana*, *strada urbana*), è stato estratto un campione casuale di 4 distributori per ogni associazione tra modalità di campagna promozionale attivata e ubicazione. Si è rilevata la quantità di carburante (in centinaia di litri) venduta nelle ultime due settimane di gennaio di 2006. Di seguito è riportata la tabella delle quantità medie:

Ubicazione	Campagna promozionale		$\bar{X}_{j..}$
	Raccolta punti	Gratta e vinci	
Autostrada	60,375	63,875	<b>62,125</b>
Extraurbana	49,075	52,175	<b>50,625</b>
Urbana	44,000	40,000	<b>42,000</b>
$\bar{X}_{.k.}$	<b>51,150</b>	<b>52,017</b>	<b>51,583</b>

Sapendo che la *devianza totale* è 1775,6533 dopo aver specificato le ipotesi necessarie, si verifichi, ad un livello di significatività  $\alpha = 0,05$ , se:

- l'ubicazione del distributore influenza significativamente la quantità media di carburante venduto;
  - la modalità di campagna promozionale adottata influenza significativamente la quantità media di carburante venduto;
  - vi è *interazione* tra ubicazione del distributore e modalità di campagna promozionale adottata.
- La seguente tabella riporta, per alcune regione italiane, alcuni risultati dell'indagine multiscope Istat sulle famiglie *Aspetti della vita quotidiana*<sup>1</sup> relativi alle abitudini alimentari. In particolare le quattro variabili considerate riportano la percentuale di persone di 3 anni e più che:  $X_1$  = hanno una colazione adeguata (non assumono soltanto tè o caffè);  $X_2$  = pranzano a casa;  $X_3$  = pranzano in un bar;  $X_4$  = hanno il pranzo come pasto principale.

Regione	$X_1$	$X_2$	$X_3$	$X_4$
Veneto	80,30	73,70	2,10	70,90
Liguria	81,40	72,70	4,00	64,30
Emilia-Romagna	81,00	72,80	2,90	66,80
Toscana	81,30	74,10	3,20	63,10
Umbria	82,90	80,30	1,30	72,50
Marche	83,90	80,00	1,10	74,10

La matrice  $D$  delle *distanze city-block* calcolata sui dati standardizzati è:

$$D = \begin{bmatrix} 0 & 4,639 & 2,614 & 3,890 & 5,319 & 6,627 \\ 4,639 & 0 & 2,025 & 1,575 & 8,154 & 9,462 \\ 2,614 & 2,025 & 0 & 1,829 & 6,785 & 8,093 \\ 3,890 & 1,575 & 1,829 & 0 & 7,324 & 8,632 \\ 5,319 & 8,154 & 6,785 & 7,324 & 0 & 1,492 \\ 6,627 & 9,462 & 8,093 & 8,632 & 1,492 & 0 \end{bmatrix}.$$

- 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.

**CONTINUA SUL RETRO**

<sup>1</sup>Fonte: Istat - Stili di Vita e Condizioni di Salute - Indagine multiscope sulle famiglie. Aspetti della vita quotidiana anno 2003.

4. Sulle 100 unità di una popolazione sono state rilevate quattro variabili quantitative. Calcolata la matrice  $R$  di correlazione tra le quattro variabili, si sono ricavate dalla stessa le componenti principali. Di seguito è riportata la matrice  $C$  di correlazione tra le variabili originarie (righe) e le componenti principali (colonne):

$$C = \begin{bmatrix} -0,3926 & 0,7491 & 0,5317 & -0,0455 \\ -0,0557 & -0,8885 & 0,4329 & -0,1418 \\ -0,9200 & 0,0178 & -0,2825 & -0,2710 \\ -0,9081 & -0,2874 & 0,0298 & 0,3029 \end{bmatrix}.$$

- 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 1,58 e 0,12 interpretandone la posizione rispetto alle variabili originarie.