

Università degli Studi di Milano-Bicocca - Facoltà di Economia  
Esame di Analisi dei Dati I  
19 settembre 2006

**NB: Commentare sempre i risultati ottenuti**

1. Metodi non gerarchici aggregativi nell'analisi dei gruppi: caratteristiche generali e criteri per la determinazione del numero  $k$  dei gruppi della partizione e dei relativi centri iniziali.
2. Un'azienda che produce condizionatori per uso domestico vuole sperimentare sui propri prodotti dei nuovi dispositivi che, a parità di consumo elettrico, dovrebbero garantire una maggiore capacità refrigerante. In particolare si vuole verificare se tre differenti dispositivi  $A$ ,  $B$  e  $C$  influenzano significativamente la capacità refrigerante (misurata in BTU/h) su due differenti modelli di condizionatore: *portatile* e *fisso*. A tale proposito, per ogni associazione tra dispositivo e modello, vengono testati 5 condizionatori. La seguente tabella riporta le capacità di refrigerazione media (BTU/h):

Dispositivo	Modello		$\bar{X}_{j..}$
	<i>Portatile</i>	<i>Fisso</i>	
$A$	11188,4	10863,0	<b>11025,7</b>
$B$	11129,8	11324,6	<b>11227,2</b>
$C$	11086,2	10902,2	<b>10,994,2</b>
$\bar{X}_{.k.}$	<b>11134,8</b>	<b>11029,933</b>	<b>11082,367</b>

- a) Dopo aver specificato il modello da utilizzare e il significato dei parametri in esso contenuti, si calcolino le stime per gli effetti specifici  $\alpha_j = \mu_j - \mu..$   $j = 1, 2, 3$  imputabili alle modalità del fattore *dispositivo* commentando opportunamente;
- b) sapendo che la *devianza nei gruppi* è 1228876,8, si verifichino ad un livello di significatività  $\alpha = 0,10$  le seguenti ipotesi:
  - b1)** il tipo di dispositivo influenza significativamente la capacità media di refrigerazione;
  - b2)** il modello di condizionatore influenza significativamente la capacità media di refrigerazione;
  - b3)** vi è *interazione* tra tipo di dispositivo e modello del condizionatore.
3. Un tour operator vuole classificare i sei villaggi turistici con cui è presente in Grecia sulla base dei risultati conseguiti nel mese di agosto 2006. Per ciascuno dei sei villaggi sono noti:  $X_1 =$  *capienza (numero di posti letto)*;  $X_2 =$  *arrivi (numero di clienti)*;  $X_3 =$  *fatturato (in euro)*;  $X_4 =$  *durata media del soggiorno di ogni cliente (in giorni)*. Di seguito è riportata la matrice dei dati standardizzati:

Villaggio	$Z_1$	$Z_2$	$Z_3$	$Z_4$
<b>A</b>	0,630	1,354	1,082	0,603
<b>B</b>	0,394	-0,640	-1,082	1,508
<b>C</b>	-1,261	-0,850	-0,541	-1,206
<b>D</b>	0,630	-0,535	-0,812	0,603
<b>E</b>	1,103	1,459	1,623	-0,302
<b>F</b>	-1,497	-0,787	-0,271	-1,206

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

Partizione	Livello di distanza
$\{A, B, C, D, E, F\}$	$0 \leq d < 0,569$
$\{(C, F), A, B, D, E\}$	$0,569 \leq d < 1,516$
$\{(B, D), (C, F), A, E\}$	$1,516 \leq d < 2,024$

- a) Si completi la successione delle partizioni e si tracci il dendrogramma riferito ai sei villaggi;
  - b) si suggerisca una opportuna partizione valutando la distanza tra le partizioni che compongono la successione ottenuta al punto precedente;
  - c) si descriva la partizione individuata al punto precedente.
4. 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.