

Università degli Studi di Milano-Bicocca - Facoltà di Economia  
Esame di Analisi dei Dati (modulo A)  
14 luglio 2004

**NB: Commentare sempre i risultati ottenuti**

- I contrasti lineari: dopo aver illustrato il contesto di applicazione nell'ambito dell'analisi della varianza ad un criterio di classificazione, si illustri lo stimatore, le sue caratteristiche e le relative verifiche d'ipotesi.
- Un'azienda produttrice di snack vuole lanciare sul mercato una "confezione merenda" che comprende il nuovo snack e una bibita. Al fine di decidere quale tra le bibite  $A$ ,  $B$ ,  $C$  e  $D$  abbinare allo snack, fornisce ad un campione-test di clienti abituali una fornitura mensile del prodotto. Per ciascuna potenziale bibita da abbinare allo snack, viene selezionato un campione casuale di 5 maschi e 5 femmine. Al termine del mese di prova i clienti esprimono una valutazione di gradimento del prodotto con un punteggio da 0 a 10. Nella seguente tabella sono riportati le valutazioni medie espresse:

Sesso	Bibita				$\bar{X}_{j..}$
	$A$	$B$	$C$	$D$	
$M$	6,6	5,2	4,8	8,2	<b>6,2</b>
$F$	5,6	3,0	7,4	7,2	<b>5,8</b>
$\bar{X}_{.k.}$	<b>6,1</b>	<b>4,1</b>	<b>6,1</b>	<b>7,7</b>	<b>6</b>

Sapendo che la devianza totale è pari a 190, dopo aver specificato il modello da utilizzare, considerando un livello di significatività  $\alpha = 0,05$ , si verifichi se:

- il fattore *sex* influenza significativamente il gradimento medio;
  - il fattore *beverage* influenza significativamente il gradimento medio;
  - vi è *interazione* tra i due fattori.
- Misure di diversità: definizione, esempi ed impieghi.
  - Per una località montana si sono rilevate le seguenti quattro variabili meteorologiche negli anni 1988-1993:  $X_1 = \text{temperatura massima (in gradi centigradi)}$ ;  $X_2 = \text{temperatura minima (in gradi centigradi)}$ ;  $X_3 = \text{mm di pioggia caduti in media in un mese}$ ;  $X_4 = \text{numero di giorni con precipitazioni nevose}$ . Di seguito è riportata la matrice dei dati:

Anno	$X_1$	$X_2$	$X_3$	$X_4$
1988	13,0	-2,9	89	4
1989	13,5	-2,7	92	0
1990	14,9	-0,6	131	0
1991	13,8	-7,2	80	14
1992	12,1	-3,6	46	2
1993	12,0	-3,0	14	0

La matrice delle *distanze city-block* tra i sei anni considerati calcolata sui dati standardizzati, considerati nello stesso ordine di presentazione della precedente tabella, è:

$$D = \begin{bmatrix} 0 & 1,4835 & 5,0028 & 5,2331 & 2,8173 & 3,8781 \\ & 0 & 3,5193 & 5,7188 & 3,4990 & 3,7580 \\ & & 0 & 8,6395 & 7,0183 & 7,2773 \\ & & & 0 & 6,8518 & 8,5228 \\ & & & & 0 & 1,6710 \\ & & & & & 0 \end{bmatrix}$$

- Tracciare il dendrogramma riferito ai sei anni avvalendosi del *metodo del legame singolo*;
  - suggerire una opportuna partizione, giustificando la scelta;
  - descrivere la partizione individuata al punto precedente.
- Si illustri analiticamente la procedura per ottenere la matrice di correlazione fra variabili e componenti principali.