

Università degli Studi di Milano-Bicocca - Facoltà di Economia
Esame di Analisi dei Dati (modulo B)

16 giugno 2005

NB : Commentare sempre i risultati ottenuti.

1. Data una tabella di contingenza riferita a tre caratteri, si dimostri che l'indipendenza completa implica l'indipendenza di un qualunque carattere dalla coppia restante.
2. Una società di ricerche è interessata al posizionamento sul mercato di quattro note marche di dentifricio (A, B, C, D) secondo le caratteristiche che vengono maggiormente riconosciute al prodotto dagli utenti quali *Freschezza*, *Economicità*, *Praticità della confezione* e *Potere sbiancante*. A tale proposito viene organizzata un'indagine chiedendo ad un campione di persone di testare per una settimana ciascuno i differenti dentifrici e di classificarli secondo la caratteristica che ritengono preponderante. La tabella di contingenza riferita alle marche di dentifricio ed alle loro caratteristiche riconosciute è la seguente:

<i>Caratteristiche riconosciute</i> <i>Marche</i>	<i>Freschezza</i>	<i>Economicità</i>	<i>Praticità conf.</i>	<i>Potere sbiancante</i>
<i>A</i>	14	14	3	4
<i>B</i>	16	0	1	4
<i>C</i>	3	30	15	2
<i>D</i>	2	0	3	20

Attraverso l'analisi delle corrispondenze si sono ottenute le due seguenti matrici le cui colonne riportano rispettivamente le proiezioni dei profili *Marche* e le proiezioni dei profili *Caratteristica riconosciuta* sugli assi principali estraibili:

$$\begin{bmatrix} -0,1353 & -0,3593 & -0,1692 \\ 0,5647 & -1,0029 & 0,1409 \\ -0,7278 & 0,3075 & 0,0626 \\ 1,1708 & 0,7305 & -0,0067 \end{bmatrix} \quad \begin{bmatrix} 0,2893 & -0,9246 & 0,0155 \\ -0,7480 & 0,1650 & -0,1005 \\ -0,4569 & 0,3716 & 0,2260 \\ 1,0947 & 0,5642 & -0,0364 \end{bmatrix}.$$

- a) Calcolare l'inerzia totale e l'inerzia di ogni profilo riga e di ogni profilo colonna. Descrivere le informazioni fornite dai risultati ottenuti.
 - b) Quanti assi principali si ritiene che sia utile mantenere nell'analisi?
3. Con l'obiettivo di voler stabilire se l'olio di oliva prodotto da una azienda agricola può essere classificato come *D.O.P.* (Denominazione di Origine Protetta), sono state rilevate le caratteristiche X_1 *acidità* (% , grammi per ogni 100 grammi), X_2 *limpidezza* (scala da 1, *trasparente* a 5, *denso*), X_3 *viscosità* (scala da 1, *meno viscoso* a 5, *più viscoso*) su un campione di 6 oli d'oliva che sono già stati classificati come *D.O.P.* (i primi 4) oppure no (gli ultimi 2). I risultati sono riportati nella seguente matrice \mathbf{X} :

$$\mathbf{X} = \begin{bmatrix} 0,008 & 1 & 5 \\ 0,010 & 2 & 4 \\ 0,011 & 3 & 3 \\ 0,019 & 4 & 3 \\ \text{---} & \text{---} & \text{---} \\ 0,020 & 4 & 3 \\ 0,022 & 3 & 3 \end{bmatrix}.$$

- a) Calcolare i punteggi discriminanti rispetto alla funzione discriminante estraibile sapendo che

$$\mathbf{W}^{-1}_{(3 \times 3)} = \begin{bmatrix} 39384 & -123.29 & -13.699 \\ -123.29 & 1.3425 & 1.2603 \\ -13.699 & 1.2603 & 1.9178 \end{bmatrix}.$$

- b) Riclassificare le unità statistiche secondo la funzione discriminante estraibile. Quante sono classificate correttamente?
 - c) Sapendo che il nuovo olio ha le seguenti caratteristiche $X_1 = 0.018$, $X_2 = 4$ e $X_3 = 3$, può venirgli assegnata l'etichetta *D.O.P.*?
4. Si consideri la tabella a doppia entrata dell'esercizio 2. Si valuti la distanza chi-quadrato:
 - a) fra il profilo della marca A e il profilo della marca D ;
 - b) fra il profilo caratteristica preponderante *Economicità* e il corrispondente profilo medio.