

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

14 luglio 2005

NB : Commentare sempre i risultati ottenuti.

1. Definizione e determinazione della funzione discriminante (lineare) nel caso di $k = 2$ gruppi.
2. Il modello probabilistico prodotto di multinomiali per una tabella di contingenza riferita a due caratteri: descrizione dello schema di campionamento e determinazione della distribuzione di probabilità.
3. È stata condotta un'indagine epidemiologica in una zona industriale di una grande città con l'obiettivo di studiare le associazioni tra settore industriale e incidenza di problemi alle vie respiratorie dei lavoratori. Le informazioni relative alle malattie alle vie respiratorie contratte nei due anni precedenti all'indagine sono state ricavate dalle cartelle cliniche dei lavoratori e i risultati sono riportati nella seguente tabella:

<i>Problemi alle vie respiratorie Settore industriale</i>	<i>Assenti</i>	<i>Lievi</i>	<i>Moderati</i>	<i>Seri</i>
<i>Meccanica</i>	8	0	3	1
<i>Siderurgia</i>	13	0	2	1
<i>Chimica</i>	1	16	4	2
<i>Metallurgia</i>	2	0	3	20

Attraverso l'analisi delle corrispondenze, che ha condotto a mantenere due assi principali, si sono ottenute le due seguenti matrici le cui colonne riportano rispettivamente le proiezioni dei profili *settore industriale* e le proiezioni dei profili *problemi alle vie respiratorie* sui due assi principali mantenuti:

$$\begin{bmatrix} -0.5690 & 0.6751 \\ -0.6891 & 0.8409 \\ 1.2259 & 0.1316 \\ -0.4137 & -0.9833 \end{bmatrix}; \quad \begin{bmatrix} -0.6714 & 0.8189 \\ 1.5065 & 0.1784 \\ 0.0591 & 0.1450 \\ -0.3625 & -1.0104 \end{bmatrix}.$$

- a) Sapendo che l'inerzia totale è 1.2178, si valuti la percentuale di inerzia totale dovuta al primo e al secondo asse principale e si commenti.
 - b) Valutare i contributi assoluti e i contributi relativi ritenuti utili per il tipo di indagine, commentando i risultati ottenuti.
 - c) Si rappresentino graficamente i risultati e si commenti opportunamente alla luce di tutte le informazioni raccolte.
4. Nella seguente tabella 100 famiglie sono state classificate secondo lo stipendio mensile (in migliaia di euro) percepito dal marito (X) e quello percepito dalla moglie (Y):

Y	1 + 2	2 + 3	3 + 5	5 + 10
X				
1 + 2	32	14	3	1
2 + 3	13	6	0	1
3 + 5	13	6	1	0
5 + 10	7	2	1	0

Si descriva il modello log-lineare di indipendenza distributiva per la tabella sopra riportata e si forniscano le stime di massima verosimiglianza dei parametri di tale modello.