

Università degli Studi di Milano-Bicocca - Facoltà di Economia  
Esame di Analisi dei Dati (modulo B)  
16 febbraio 2005

NB : Commentare sempre i risultati ottenuti.

1. In una città sono stati osservati giornalmente la *condizione meteorologica* e il *livello di traffico automobilistico* per un periodo di tempo di un anno. Si è potuta costruire così la tabella delle associazioni fra i due carteri:

LIV. TRAFFICO TEMPO	<i>basso</i>	<i>medio</i>	<i>alto</i>
<i>sereno</i>	84	26	11
<i>variabile</i>	29	98	29
<i>pioggia</i>	7	26	55

Attraverso l'analisi delle corrispondenze si sono ottenute le seguenti matrici, le cui colonne riportano rispettivamente le proiezioni dei profili riga e le proiezioni dei profili colonna sugli assi principali estraibili:

$$\begin{bmatrix} -0,7566 & 0,1937 \\ 0,1785 & -0,4078 \\ 0,7238 & 0,4567 \end{bmatrix} \quad \begin{bmatrix} -0,7737 & 0,1740 \\ 0,1933 & -0,4202 \\ 0,6721 & 0,4438 \end{bmatrix}.$$

- a) Quanti assi principali si ritiene che sia utile mantenere nell'analisi?  
b) Calcolare l'inerzia totale e l'inerzia di ogni profilo riga e di ogni profilo colonna. Descrivere le informazioni fornite dai risultati ottenuti.  
c) In base alle risposte ai punti a) e b) si effettui un'opportuna rappresentazione grafica dei risultati e si commenti adeguatamente.
2. Il proprietario di un cinema multisala appena inaugurato è interessato a verificare il grado di soddisfazione degli utenti rispetto agli spettacoli proposti e all'arredamento delle sale. A tal fine predispose un'intervista presso un campione di 100 spettatori. I risultati rispetto al gradimento del film e al comfort della sala sono riportati nella seguente tabella:

<i>Gradimento film</i> <i>Comfort sala</i>	<i>poco</i>	<i>abbastanza</i>	<i>molto</i>
<i>scarso</i>	2	1	2
<i>sufficiente</i>	4	5	6
<i>buono</i>	14	16	15
<i>ottimo</i>	10	13	12

Si descriva il modello log-lineare di distribuzioni di riga uniformi per la tabella sopra riportata, commentando adeguatamente rispetto al contesto di riferimento. Si forniscano le stime di massima verosimiglianza delle frequenze attese e dei parametri di tale modello.

3. Si consideri la tabella a doppia entrata del precedente esercizio e si valuti la distanza chi-quadrato:
- a) fra i profili del giudizio sul comfort della sala *sufficiente* e *buono*;  
b) fra il profilo gradimento del film *molto* e il corrispondente profilo medio.
4. Si illustri l'origine della statistica test impiegata per effettuare verifiche d'ipotesi sull'adattamento di un modello log-lineare per tabelle di contingenza a doppia entrata e le modalità per selezionare un modello secondo il criterio gerarchico.
5. Si descrivano gli obiettivi dell'analisi discriminante e dopo aver dato la definizione di funzione discriminante si esponcano i criteri per il numero di funzioni discriminanti utili ai fini dell'indagine.