

**Università degli Studi di Milano-Bicocca – Facoltà di Economia**  
**Esame di Analisi dei Dati (modulo B)**  
 28 giugno 2004

**N.B.: Commentare sempre i risultati ottenuti**

1) In un piccolo comune italiano si è verificato da qualche anno un cambio di amministrazione, con la vittoria alle elezioni di una lista civica. Una società che si occupa di sondaggi elettorali è interessata a valutare come sia modificato il giudizio circa l'operato della giunta a seguito del cambiamento di amministrazione. Un campione di 100 elettori è quindi chiamato a fornire un giudizio su come il comune era amministrato *prima* e su come è amministrato *dopo* la vittoria della lista civica. Si riporta di seguito la tabella delle associazioni fra le due variabili:

<b>PRIMA \ DOPO</b>	<i>pessimo</i>	<i>sufficiente</i>	<i>buono</i>	<i>ottimo</i>
<i>pessimo</i>	1	5	6	15
<i>sufficiente</i>	7	8	6	9
<i>buono</i>	6	8	5	4
<i>ottimo</i>	9	4	2	5

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

$$\begin{bmatrix} -0.5775 & -0.0698 & -0.0029 \\ 0.0384 & 0.0535 & 0.0089 \\ 0.2347 & 0.2551 & -0.0060 \\ 0.4521 & -0.2794 & -0.0027 \end{bmatrix} \quad \begin{bmatrix} 0.5921 & -0.1632 & 0.0006 \\ 0.1165 & 0.2215 & -0.0078 \\ -0.1604 & 0.1800 & 0.0085 \\ -0.4086 & -0.1577 & -0.0037 \end{bmatrix}$$

- Dopo aver stabilito quanti assi principali è utile mantenere nell'analisi, si effettui un'opportuna rappresentazione grafica dei risultati dell'analisi delle corrispondenze sopra riportati, commentando adeguatamente.
- Si valuti la qualità della rappresentazione dei profili riga nello spazio degli assi principali mantenuti.
- Si calcoli l'inerzia di ogni profilo riga, commentando i risultati ottenuti.

2) Modelli probabilistici per tabelle di contingenza.

3) La matrice dei punteggi discriminanti: definizione e impieghi.

4) Con l'obiettivo di spiegare l'appartenenza di un'automobile ad una delle 4 categorie indicate con A, B, C e D, sono state rilevate, su un campione di 12 automobili (di cui 4 appartengono alla categoria A, 3 alla categoria B, 3 alla categoria C e 2 alla categoria D), le seguenti caratteristiche:  $X_1 = \text{Cilindrata}$ ;  $X_2 = \text{Numero di cavalli}$ ;  $X_3 = \text{Tempo di accelerazione da 0 a 100 km/h (in secondi)}$ ;  $X_4 = \text{Consumo (in km per litro)}$ .

I vettori dei centri dei punteggi discriminanti rispetto alle funzioni discriminanti estraibili sono di seguito riportati:

$$c_1 = \begin{bmatrix} -13.7098 \\ -16.7922 \\ -9.9753 \\ -15.3382 \end{bmatrix} \quad c_2 = \begin{bmatrix} 12.5618 \\ 13.2601 \\ 13.55914 \\ 14.4169 \end{bmatrix} \quad c_3 = \begin{bmatrix} 17.2648 \\ 17.8704 \\ 17.6983 \\ 17.1314 \end{bmatrix}$$

Considerato che gli autovalori della matrice  $W^{-1}B$  e la matrice  $A$  dei corrispondenti autovettori normalizzati sono rispettivamente:

$$\lambda_1 = 13.8815; \lambda_2 = 0.6034; \lambda_3 = 0.083; \quad e \quad A = \begin{bmatrix} 0.0009 & 0.0010 & 0.0014 \\ -0.0049 & 0.0220 & 0.0376 \\ -0.7866 & 0.9973 & -0.0383 \\ -0.6174 & -0.0701 & 0.9986 \end{bmatrix}$$

- Si determini il numero di funzioni discriminanti da mantenere nell'analisi, sia dal punto di vista descrittivo che inferenziale, specificando le adeguate ipotesi da assumere e adottando un livello di significatività  $\alpha = 0.05$ .
- Si classifichi opportunamente in uno dei quattro gruppi una nuova automobile, caratterizzata da una cilindrata pari a 1400, un numero di cavalli pari a 240, un tempo di accelerazione da 0 a 100 km/h pari a 10 secondi e un consumo di 12.2 km/litro.

5) Si illustrino gli obiettivi generali dei modelli log-lineari e si presenti in dettaglio il modello log-lineare saturo per una tabella di contingenza  $r \times c$ .