

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

5 aprile 2005

NB : Commentare sempre i risultati ottenuti.

1. Illustrare il tipo di dati di cui bisogna disporre per applicare l'analisi discriminante e descrivere gli obiettivi dell'analisi stessa.
2. Si consideri una tabella a triplice entrata riferita alla distribuzione di tre caratteri qualitativi e si illustrino le diverse distribuzioni di frequenze che si possono definire.
3. Illustrare gli indici sintetici utilizzati nell'analisi delle corrispondenze, fornendo espressione analitica, significato ed impiego.
4. Una nota azienda di attrezzatura sportiva dedicata all'atletica leggera è interessata a lanciare un nuovo tipo di giavellotto. A tal proposito l'azienda deve valutare 4 diversi prototipi, *A*, *B*, *C* e *D*, fornitigli dal reparto 'Ricerca e Sviluppo Nuovi Materiali'. Con questo obiettivo viene chiesto ad un campione di 25 atleti di testare, in tutta sicurezza, le prestazioni dei prototipi e di classificare ciascuno di essi secondo la caratteristica che ritengono preponderante tra *Elasticità*, *Leggerezza*, *Manegevolezza* e *Velocità*. La tabella di contingenza riferita ai prototipi ed alla loro classificazione secondo la caratteristica preponderante è la seguente:

<i>Caratteristica preponderante Prototipi</i>	<i>Elasticità</i>	<i>Leggerezza</i>	<i>Manegevolezza</i>	<i>Velocità</i>
<i>A</i>	14	0	7	4
<i>B</i>	16	0	5	4
<i>C</i>	3	14	6	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 *Prototipi* e le proiezioni dei profili *Caratteristica preponderante* sugli assi principali estraibili:

$$\begin{bmatrix} -0,2720 & -0,5189 & 0,1069 \\ -0,3138 & -0,5809 & -0,1008 \\ 1,2383 & 0,2085 & -0,0035 \\ -0,6525 & 0,8913 & -0,0026 \end{bmatrix} \quad \begin{bmatrix} -0,2512 & -0,6728 & -0,0512 \\ 1,6963 & 0,3469 & -0,0471 \\ 0,1304 & -0,2070 & 0,1397 \\ -0,5898 & 0,7679 & -0,0161 \end{bmatrix}.$$

- a) Quanti assi principali si ritiene che sia utile mantenere nell'analisi?
 - b) In base alla risposta al punto a) si effettui una opportuna rappresentazione grafica dei risultati e si commenti adeguatamente.
 - c) Valutare la qualità della rappresentazione dei profili riga nello spazio degli assi principali mantenuti.
5. Nell'ambito del progetto *Comuni Ricicloni 2005* sponsorizzato da Legambiente è stato classificato un campione di 90 comuni della regione Lombardia nelle 4 fasce di merito in ordine decrescente; rispettivamente 26 comuni sono risultati di prima fascia, 20 di seconda, 28 di terza e 16 di quarta fascia. Le caratteristiche che hanno portato a questa classificazione sono state: *percentuale di raccolta differenziata attivata*, *produzione pro-capite di rifiuti urbani* (kg/ab/anno), *costo pro-capite annuo della gestione dei rifiuti urbani* (euro/ab/anno). I vettori dei centri dei punteggi discriminanti rispetto alle funzioni discriminanti estraibili sono di seguito riportati:

$$\mathbf{c}_1 = \begin{bmatrix} +1,108 \\ -0,021 \\ -0,710 \\ -0,531 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} +0,141 \\ -0,438 \\ +0,303 \\ -0,212 \end{bmatrix}, \quad \mathbf{c}_3 = \begin{bmatrix} -0,006 \\ +0,032 \\ +0,014 \\ -0,055 \end{bmatrix}.$$

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

$$\lambda_1 = 0,6165 \quad \lambda_2 = 0,0932 \quad \lambda_3 = 0,0009,$$

$$\mathbf{A}_{(3 \times 3)} = \begin{bmatrix} 0,232 & 0,081 & 0,063 \\ -4,593 & 1,864 & 3,707 \\ 0,014 & 0,069 & -0,030 \end{bmatrix};$$

si stabiliscano:

- a) il numero di funzioni discriminanti da mantenere nell'analisi dal punto di vista inferenziale;
- b) tenendo conto della scelta effettuata al punto precedente, a quale fascia potrebbe appartenere un ulteriore comune che presenta i seguenti valori delle variabili in esame: 53, 1.25, 57.