

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

4 febbraio 2004

NB: Commentare sempre i risultati ottenuti.

1. Modelli log-lineari per tabelle di contingenza a doppia entrata di dimensione $r \times c$:
 - a) spiegare il significato dell'ipotesi nulla $H_0 : \theta_j^B = 0$ per $j = 1, \dots, c$ e presentare un modo equivalente di esprimere la stessa;
 - b) illustrare la costruzione della statistica-test per verificare la precedente ipotesi nulla contro l'alternativa del modello saturo, fornendo significato e formalizzazione dell'ipotesi alternativa.

2. Obiettivi generali dell'analisi discriminante, definizione e determinazione della prima funzione discriminante.

3. Con l'obiettivo di spiegare il proprio rischio di credito una banca classifica 7 aziende, sue clienti, in base al fatto se in passato queste siano risultate insolventi o meno, in particolare 3 sono state classificate nel gruppo '*insolventi in passato*' e le restanti nel gruppo '*solventi in passato*'. A questo scopo sulle aziende considerate vengono rilevate le seguenti variabili: $X_1 =$ numero di prestiti richiesti negli ultimi tre anni, $X_2 =$ numero di conti correnti aperti presso la banca. I risultati sono riportati nella matrice \mathbf{X} seguente:

$$\mathbf{X}_{(7 \times 2)} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 4 & 1 \\ \hline 3 & 3 \\ 0 & 4 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

- a) Calcolare la matrice \mathbf{T} di devianze e codevianze di \mathbf{X} .
- b) Scomporre la matrice \mathbf{T} in somma delle matrici di devianze e codevianze nei gruppi e devianze e codevianze fra i gruppi.
- c) Calcolare i punteggi discriminanti rispetto alla funzione discriminante estraibile sapendo che

$$\mathbf{W}^{-1}_{(2 \times 2)} = \begin{bmatrix} 0, \overline{2} & 0, \overline{16} \\ 0, \overline{16} & 0, \overline{375} \end{bmatrix}.$$

- d) Riclassificare le unità statistiche secondo la funzione discriminante estraibile. Quante sono classificate correttamente?

4. La distanza chi-quadrato: definizione, proprietà e impieghi.

5. Una azienda sanitaria è interessata a valutare la possibilità di ristrutturare alcuni degli ambienti dove vengono offerti i servizi da alcuni ambulatori. A tal proposito l'azienda sanitaria decide di svolgere un'indagine su un campione di 729 pazienti. Gli usufruttori degli ambulatori coinvolti nell'indagine sono classificati secondo l'*ambulatorio di riferimento* (A, B, C, D) e secondo il *giudizio* espresso circa il grado di comfort degli spazi di attesa prima di una visita (*per niente confortevole, poco confortevole, abbastanza confortevole, molto confortevole*). La tabella di contingenza riferita alla classificazione dei pazienti è la seguente:

<i>Ambulatorio</i>	<i>Giudizio sul comfort</i>	<i>per niente confort.</i>	<i>poco confort.</i>	<i>abbastanza confort.</i>	<i>molto confort.</i>
<i>A</i>		56	1	20	9
<i>B</i>		4	10	12	4
<i>C</i>		23	20	69	46
<i>D</i>		13	14	91	337

Attraverso l'analisi delle corrispondenze si sono ottenute le due seguenti matrici le cui colonne riportano rispettivamente le proiezioni dei profili *Ambulatorio* e le proiezioni dei profili *Giudizio sul comfort* sugli assi principali estraibili:

$$\begin{bmatrix} 1,4937 & -0,4636 & 0,0363 \\ 0,4068 & 1,1231 & 0,3852 \\ 0,2690 & 0,4556 & -0,1440 \\ -0,4026 & -0,1446 & 0,0177 \end{bmatrix} \quad \begin{bmatrix} 1,4441 & -0,3645 & 0,0492 \\ 0,1896 & 1,0786 & 0,2670 \\ 0,1397 & 0,3183 & -0,1479 \\ -0,4394 & -0,1885 & 0,0295 \end{bmatrix}.$$

- 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 e colonna nello spazio degli assi principali mantenuti.