

Università degli Studi di Milano-Bicocca - Facoltà di Economia
Esame di Analisi dei Dati (modulo B)
 25 febbraio 2004

NB: Commentare sempre i risultati ottenuti.

1. Si dimostri che la matrice di devianze e codevianze \mathbf{T} , riferita a p variabili quantitative, si scompone, qualora i dati lo consentano, nella somma di due matrici.
2. Si descrivano i diversi tipi di distribuzioni di frequenze che si possono definire in una tabella a triplice entrata riferita a caratteri di natura qualitativa e si illustrino le principali forme di indipendenza.
3. Data una tabella di contingenza di dimensione $r \times c$, con $c < r$, si definisca e si determini il primo asse principale nello spazio dei profili riga o colonna che si ritiene più adeguato. Si estenda la definizione ai successivi assi, stabilendo il numero massimo di assi principali estraibili.
4. Nell'ambito di uno studio sulla sicurezza del lavoro sono stati classificati 250 infortuni occorsi ai lavoratori di un gruppo di aziende di un certo settore nel corso di un anno. Il campione rilevato è stato classificato secondo il *tipo di infortunio*, T , (*frattura, ferita, contusione, altro*) e la *parte del corpo* oggetto dell'infortunio, P , (*gambe, braccia, altro*). I risultati sono riportati nella seguente tabella:

| T | P | <i>gambe</i> | <i>braccia</i> | <i>altro</i> |
|-------------------|-----|--------------|----------------|--------------|
| <i>frattura</i> | | 22 | 32 | 10 |
| <i>ferita</i> | | 21 | 24 | 15 |
| <i>contusione</i> | | 19 | 33 | 13 |
| <i>altro</i> | | 18 | 27 | 16 |

Fissando in $\alpha = 0,05$ il livello di significatività, stabilire quale modello log-lineare gerarchico è idoneo a descrivere la relazione fra *tipo di infortunio* e *parte del corpo* oggetto dell'infortunio nella popolazione corrispondente.

5. Con l'obiettivo di spiegare l'appartenenza di una birra ad un certo tipo, si rilevano presso un campione di 69 birre, delle quali 13 di tipo *trappista*, 17 di tipo *abbazia*, 10 di tipo *ale*, 16 di tipo *stout* e 13 di tipo *lager*, le caratteristiche: *prezzo* (in euro), *calorie* (in kcal.), *gradazione alcolica*.

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

$$\mathbf{c}_1 = \begin{bmatrix} +0,757 \\ +2,191 \\ +0,949 \\ -0,394 \\ -3,866 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} +0,468 \\ -0,821 \\ -0,572 \\ +1,511 \\ -0,814 \end{bmatrix}, \quad \mathbf{c}_3 = \begin{bmatrix} -0,139 \\ +0,004 \\ +0,100 \\ +0,056 \\ -0,013 \end{bmatrix}.$$

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

$$\lambda_1 = 4,6070 \quad \lambda_2 = 0,9795 \quad \lambda_3 = 0,0063,$$

$$\mathbf{A}_{(3 \times 3)} = \begin{bmatrix} 0,411 & -0,965 & 0,259 \\ 0,075 & 0,018 & -0,066 \\ -0,186 & 0,287 & 1,320 \end{bmatrix};$$

si stabiliscano:

- a) il numero di funzioni discriminanti da mantenere nell'analisi sia dal punto di vista descrittivo che inferenziale;
- b) a quale tipo potrebbe essere assegnata una nuova birra che presenta i seguenti valori delle variabili in esame: 5; 155; 5.3.