

Università degli Studi di Milano-Bicocca - Facoltà di Economia
Esame di Analisi dei Dati (modulo II)
 13 luglio 2006

NB : Commentare sempre i risultati ottenuti.

1. I responsabili di una grande palestra sono interessati a verificare il livello di soddisfazione dei propri utenti rispetto alla *Preparazione* e *Assistenza* dei propri istruttori. A tal proposito viene organizzata un'indagine presso un campione di 174 utenti abituali a cui viene chiesto di esprimere un giudizio in termini di livello di soddisfazione: *insoddisfatto (I)*, *soddisfatto (S)*, *più che soddisfatto (PS)*, *estremamente soddisfatto (ES)*. La tabella di contingenza contenente i dati risultanti dall'indagine è la seguente:

| <i>soddisfaz. Assistenza</i> | <i>I</i> | <i>S</i> | <i>PS</i> | <i>ES</i> |
|---|----------|----------|-----------|-----------|
| <i>soddisfaz. Preparazione istruttori</i> | | | | |
| <i>I</i> | 3 | 6 | 4 | 5 |
| <i>S</i> | 4 | 8 | 5 | 2 |
| <i>PS</i> | 11 | 20 | 27 | 15 |
| <i>ES</i> | 12 | 8 | 12 | 32 |

Attraverso l'analisi delle corrispondenze si sono ottenute le due seguenti matrici le cui colonne riportano rispettivamente le proiezioni dei profili *soddisfazione per la Preparazione degli istruttori* e le proiezioni dei profili *soddisfazione per l'Assistenza agli utenti* sugli assi principali estraibili:

$$\begin{bmatrix} 0,1098 & 0,1773 & -0,0746 \\ 0,4640 & 0,2465 & 0,0469 \\ 0,2456 & -0,1195 & -0,0013 \\ -0,4488 & 0,0133 & 0,0085 \end{bmatrix} \quad \begin{bmatrix} -0,0471 & 0,0957 & 0,0596 \\ 0,3847 & 0,1418 & -0,0241 \\ 0,2368 & -0,1857 & 0,0029 \\ -0,4835 & 0,0017 & -0,0169 \end{bmatrix}.$$

- a) Calcolare l'inerzia totale e l'inerzia di ogni profilo riga e di ogni profilo colonna. Descrivere le informazioni fornite dai risultati ottenuti.
- b) Sfruttando le informazioni del punto precedente, misurare la connessione tra i due caratteri.
- c) Quanti assi principali si ritiene che sia utile mantenere nell'analisi?
2. Un'agenzia di assicurazioni è interessata a valutare le performance dei propri agenti in vista di una promozione da assegnare ad uno di essi. A tal proposito classifica i suoi 9 dipendenti in *'efficienti'* e *'non efficienti'* in relazione alla loro performance rilevata negli ultimi sei mesi. In particolare 5 dei 9 dipendenti sono giudicati *'efficienti'*. I dipendenti vengono classificati in base alle seguenti caratteristiche: *età in anni compiuti*, *anzianità di esperienza professionale* (in anni) e *numero medio di contatti per cliente* prima della stipulazione della polizza. La matrice dei dati \mathbf{X} e l'autovettore \mathbf{a} associato all'autovalore della matrice $\mathbf{W}^{-1}\mathbf{B}$ sono rispettivamente:

$$\mathbf{X}_{(9 \times 3)} = \begin{bmatrix} 60 & 9 & 1 \\ 58 & 6 & 3 \\ 45 & 8 & 2 \\ 37 & 7 & 2 \\ 55 & 12 & 4 \\ \text{---} & \text{---} & \text{---} \\ 35 & 4 & 2 \\ 30 & 3 & 4 \\ 45 & 10 & 3 \\ 47 & 6 & 4 \end{bmatrix} \quad \mathbf{a}_{(3 \times 1)} = \begin{bmatrix} 0.0730 \\ 0.1346 \\ -0.5776 \end{bmatrix}.$$

- a) Commentare il contributo delle singole caratteristiche al punteggio della funzione discriminante estraibile.
- b) Calcolare i punteggi discriminanti rispetto alla funzione discriminante estraibile.
- c) Riclassificare le unità statistiche secondo la funzione discriminante estraibile. Quante sono classificate correttamente?
3. In riferimento ad una distribuzione di frequenze congiunte di tre caratteri, si dimostri che l'indipendenza di un carattere dalla coppia restante implica l'indipendenza marginale fra opportuni caratteri.
4. Modelli log-lineari per tabelle di contingenza a doppia entrata di dimensione $r \times c$:
- a) spiegare il significato dell'ipotesi nulla $H_0 : \theta_i^A = 0$ per $i = 1, \dots, r$ 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.