

Università degli Studi di Milano-Bicocca - Facoltà di Economia
Esame di Analisi dei Dati (modulo B)
 23 settembre 2005

NB : Commentare sempre i risultati ottenuti.

1. Illustrare la costruzione della statistica-test per verificare l'ipotesi nulla $H_0 = \{ \text{la distribuzione congiunta della tabella di contingenza bivariata è uniforme} \}$ contro l'alternativa $H_1 = \{ \text{modello saturo} \}$.
2. Si consideri la seguente distribuzione congiunta di frequenze assolute riferita ai caratteri X_1 numero di case possedute (1, 2, più di due), X_2 frequenza mensile al ristorante o affini (1, più di uno), X_3 reddito familiare (basso, alto), rilevati su un campione di 50 soggetti:

		X_3			
		basso		alto	
X_2		1	> 1	1	> 1
X_1					
1		12	9	6	3
2		5	1	4	5
> 2		1	1	2	1

Si verifichi che X_1 e X_2 sono indipendenti marginalmente, ma non condizionatamente a X_3 commentando opportunamente.

3. Una società di assicurazione vuole valutare se un cliente che stipula una polizza *RCA* avrà 0, 1, 2 o più incidenti nel prossimo anno. Le informazioni su cui basa questa ricerca si riferiscono a 8 soggetti, clienti della società da almeno due anni e sono relative alle valutazioni anagrafiche del cliente (X_1 età dell'assicurato), alla sua storia assicurativa (X_2 attestato di rischio) e alle caratteristiche dell'auto posseduta (X_3 cilindrata del motore, in cm^3). I clienti osservati sono stati così classificati: nell'ultimo anno hanno avuto 0 incidenti due di loro, 1 incidente altri tre di loro ed infine gli ultimi tre hanno avuto 2 o più incidenti. Ai fini dell'obiettivo dell'indagine è stata effettuata sui dati disponibili, dopo averli standardizzati opportunamente, un'analisi discriminante lineare. Di seguito sono riportate le matrici \mathbf{X} dei dati standardizzati e \mathbf{Y} dei punteggi discriminanti (non standardizzati) rispetto alle due funzioni discriminanti estraibili:

$$\mathbf{X}_{(8 \times 3)} = \begin{bmatrix} -1,228 & 2,098 & -1,216 \\ -1,520 & 0,745 & -1,502 \\ \text{-----} \\ -0,526 & -0,609 & 0,215 \\ 0,058 & -0,609 & -0,358 \\ 0,351 & 0,203 & -0,072 \\ \text{-----} \\ 1,111 & -0,880 & 0,787 \\ 0,935 & -0,609 & 1,359 \\ 0,818 & -0,338 & 0,787 \end{bmatrix} \quad \mathbf{Y}_{(8 \times 2)} = \begin{bmatrix} -12,764 & 1,575 \\ -12,320 & -0,877 \\ \text{-----} \\ -0,128 & -0,714 \\ 0,140 & -1,370 \\ 0,710 & 0,255 \\ \text{-----} \\ 8,569 & -0,259 \\ 9,352 & 0,850 \\ 6,442 & 0,541 \end{bmatrix}$$

- a) Si calcolino le correlazioni tra le variabili originarie e le funzioni discriminanti e si commentino i risultati ottenuti.
 - b) Utilizzando tutte le informazioni disponibili si effettui una rappresentazione grafica dell'analisi interpretando opportunamente.
4. Una azienda è interessata a incentivare le motivazioni al lavoro dei propri dipendenti. A tal proposito classifica un campione di impiegati di un ufficio secondo la loro *Produttività* (nulla, bassa, media, alta) e il loro *Grado di Soddisfazione sul lavoro* (nullo, basso, medio, alto).

Attraverso l'analisi delle corrispondenze si sono ottenute le due seguenti matrici le cui colonne riportano rispettivamente le proiezioni dei profili *Produttività* e le proiezioni dei profili *Grado di soddisfazione* sugli assi principali estraibili:

$$\begin{bmatrix} 0,1065 & -0,2929 & -0,1883 \\ -0,3161 & -0,2078 & 0,0724 \\ -0,3875 & 0,1141 & -0,0180 \\ 0,5439 & 0,0339 & 0,0139 \end{bmatrix} \quad \begin{bmatrix} -0,3170 & 0,0993 & 0,2178 \\ -0,1317 & -0,1733 & -0,0047 \\ -0,3254 & 0,1378 & -0,0351 \\ 0,7724 & 0,0487 & 0,0009 \end{bmatrix}.$$

Sapendo che si è ritenuto utile mantenere nell'analisi due assi principali, valutare la qualità della rappresentazione dei diversi livelli di *Produttività* nello spazio degli assi principali mantenuti.