

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

19 settembre 2006

NB : Commentare sempre i risultati ottenuti.

1. Una compagnia aerea è interessata a valutare il livello di soddisfazione dei propri clienti e il loro grado di fidelizzazione. A tal proposito somministra un questionario a 7 passeggeri dei propri voli e li classifica in due gruppi: coloro che dichiarano che 'riacquisteranno' un volo con la stessa compagnia aerea e coloro che 'non riacquisteranno' un volo con la stessa compagnia aerea. In particolare, dei 7 passeggeri, 4 sono classificati nel primo gruppo e 3 nel secondo. Le domande considerate nel questionario sono riferite a: *cortesìa e disponibilità del personale (di terra e di volo)*, *informazioni chiare ed esaurienti (annunci a terra, in volo, sito web)* e *adeguatezza del prezzo del biglietto al servizio offerto*. Ogni domanda viene codificata con un punteggio da 1 (per niente soddisfatto) a 5 (molto soddisfatto). La matrice dei dati \mathbf{X} partizionata nei due gruppi 'riacquisteranno' e 'non riacquisteranno' e la matrice \mathbf{W}^{-1} sono rispettivamente:

$$\mathbf{X}_{(7 \times 3)} = \begin{bmatrix} 4 & 4 & 4 \\ 3 & 4 & 4 \\ 4 & 5 & 4 \\ 5 & 5 & 3 \\ \hline 2 & 3 & 2 \\ 2 & 1 & 3 \\ 1 & 2 & 1 \end{bmatrix} \quad \mathbf{W}^{-1}_{(3 \times 3)} = \begin{bmatrix} 0,4528 & -0,2076 & -0,1132 \\ -0,2076 & 0,5535 & 0,3019 \\ -0,1132 & 0,3019 & 0,5283 \end{bmatrix}.$$

- a) Si applichi il test per verificare la significatività della funzione discriminante estraibile.
b) Si calcolino i punteggi discriminanti rispetto alla funzione discriminante estraibile.
c) Si commenti la funzione discriminante estraibile in base ai punteggi calcolati ai punti precedenti.
d) Sapendo che un nuovo cliente ha assegnato alle domande rispettivamente i punteggi 4, 3, 5, si può ritenere che riacquisterà il volo con la compagnia aerea o no?
2. Una società di assicurazioni effettua un'indagine sul numero di incidenti sul lavoro in una data città. Durante il primo semestre del 2006 (181 giorni) registra il *livello di incidentalità* \mathbf{L} (*basso* - un incidente sul lavoro per giorno; *medio* - due o tre incidenti al giorno; *alto* - più di tre incidenti al giorno) e *tipologia del giorno che precede l'incidente* \mathbf{T} (*lavorativo, non lavorativo*). La seguente tabella riporta i dati, relativi ai 181 giorni:

\mathbf{T}	\mathbf{L}	
	<i>lav.</i>	<i>non lav.</i>
<i>basso</i>	58	52
<i>medio</i>	26	25
<i>alto</i>	7	13

Volendo verificare se il livello di incidentalità è influenzato significativamente dalla tipologia del giorno che precede la rilevazione dell'incidente sul lavoro, fissando $\alpha = 0,05$, stabilire quale modello log-lineare gerarchico sia idoneo a descrivere la relazione fra i due caratteri nella popolazione corrispondente.

3. Dopo aver descritto la natura dei dati di cui bisogna disporre per applicare l'analisi delle corrispondenze ed aver spiegato gli obiettivi dell'analisi stessa ricorrendo eventualmente a qualche esempio, si definisca il primo asse principale nello spazio dei profili riga (o colonna), illustrandone il significato geometrico e la soluzione analitica.
4. 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.