

**Università degli Studi di Milano-Bicocca - Facoltà di Economia**  
**Esame di Analisi dei Dati (modulo II)**  
 12 giugno 2006

*NB : Commentare sempre i risultati ottenuti.*

1. Dopo aver illustrato il contesto applicativo dell'analisi discriminante si dia la definizione di funzione discriminante e se ne ricavi la media aritmetica e la varianza.
2. Si descriva il modello probabilistico multipoissoniano per una tabella di contingenza a doppia entrata.
3. Una società di ricerche è interessata al posizionamento sul mercato di quattro note marche di birra (*A, B, C, D*) secondo le caratteristiche che vengono maggiormente riconosciute al prodotto dagli utenti quali *Economicità, Contenuto calorico, Contenuto alcolico* e *Tipo di confezione*. A tale proposito viene organizzata un'indagine campionaria nella quale a ciascun intervistato viene chiesto di attribuire ad ognuna delle 4 marche la caratteristica ritenuta preponderante. I risultati sintetici relativi all'Analisi delle Corrispondenze applicata sono di seguito riportati:

*autovalori*  $\lambda_1 = 0,1864$   $\lambda_2 = 0,0584$   $\lambda_3 = 0,0099$

| Asse                      | Contributi ASSOLUTI |        |        | INERZIA | Contributi RELATIVI |        |        |
|---------------------------|---------------------|--------|--------|---------|---------------------|--------|--------|
|                           | 1°                  | 2°     | 3°     |         | 1°                  | 2°     | 3°     |
| <b>Profili Riga</b>       |                     |        |        |         |                     |        |        |
| <i>marca A</i>            | 0,2856              | 0,4377 | 0,0267 | 0,3105  | 0,6735              | 0,3232 | 0,0033 |
| <i>marca B</i>            | 0,4601              | 0,0005 | 0,1394 | 0,3424  | 0,9839              | 0,0003 | 0,0157 |
| <i>marca C</i>            | 0,0449              | 0,0248 | 0,8304 | 0,0706  | 0,4651              | 0,0803 | 0,4546 |
| <i>marca D</i>            | 0,2094              | 0,5371 | 0,0035 | 0,2765  | 0,5543              | 0,4452 | 0,0005 |
| <b>Profili Colonna</b>    |                     |        |        |         |                     |        |        |
| <i>Economicità</i>        | 0,4669              | 0,1980 | 0,0001 | 0,3872  | 0,8828              | 0,1172 | 0,0000 |
| <i>Contenuto calorico</i> | 0,3500              | 0,0408 | 0,3042 | 0,2774  | 0,9239              | 0,0337 | 0,0424 |
| <i>Contenuto alcolico</i> | 0,1226              | 0,0068 | 0,6555 | 0,1167  | 0,7693              | 0,0134 | 0,2173 |
| <i>Tipo di confezione</i> | 0,0604              | 0,7544 | 0,0402 | 0,2187  | 0,2023              | 0,7906 | 0,0071 |

- a) Individuare quale *marca di birra* e quale *caratteristica* risultano più importanti alla fini della connessione tra i due caratteri.
  - b) Valutare se si può considerare ragionevole mantenere nell'analisi 2 assi principali.
  - c) Fissata l'attenzione sullo spazio degli assi principali nei quali proiettare i profili ritenuti utili per l'obiettivo dell'indagine:
    - c1) fornire indicazioni su come interpretare gli assi principali mantenuti;
    - c2) valutare la qualità della rappresentazione dei profili in relazione al numero di assi principali mantenuti.
4. Una regione montana è interessata a migliorare la qualità dell'offerta per quel che riguarda il turismo invernale. A tale scopo predispone un'indagine che rileva alcune caratteristiche di 8 complessi turistici esistenti sul proprio territorio, dei quali 3 sono catalogati a categoria "3 stelle", 2 sono catalogati a categoria "4 stelle" e 3 a "5 stelle" secondo la classificazione assegnata dall'Ente Regionale del Turismo. Per ogni complesso turistico si osservano il *numero totale di km. di piste* dedicate allo sci alpino, il *numero di impianti skylift* esistenti ed il *numero di posti letto totali* offerti dal complesso (in hotel, residence, ecc.).
- Ai fini dell'obiettivo dell'indagine è stata applicata sui dati rilevati l'Analisi Discriminante Lineare. È noto che gli autovalori della matrice  $\mathbf{W}^{-1}\mathbf{B}$  sono risultati  $\lambda_1 = 44,3556$  e  $\lambda_2 = 5,1394$ . Di seguito sono riportate le matrici  $\mathbf{Z}$  dei dati standardizzati e  $\mathbf{Y}$  dei punteggi standardizzati rispetto alle due funzioni discriminanti estraibili:

$$\mathbf{Z}_{(8 \times 3)} = \begin{bmatrix} -0,8843 & -0,9375 & -1,4097 \\ -0,7438 & -0,5625 & -1,1063 \\ -1,0249 & -0,8125 & -1,0051 \\ \text{---} & \text{---} & \text{---} \\ -0,5564 & -0,1875 & 0,4109 \\ -0,1816 & -0,4375 & 0,6132 \\ \text{---} & \text{---} & \text{---} \\ 0,6618 & 0,0625 & 0,5626 \\ 1,5988 & 0,8125 & 0,8155 \\ 1,1303 & 2,0625 & 1,1189 \end{bmatrix} \quad \mathbf{Y}_{(8 \times 2)} = \begin{bmatrix} -1,4111 & -0,3269 \\ -1,1589 & -0,2647 \\ -1,0052 & 0,3775 \\ \text{---} & \text{---} \\ 0,4575 & 1,4286 \\ 0,7876 & 1,3032 \\ \text{---} & \text{---} \\ 0,6770 & -0,1587 \\ 0,8479 & -1,4378 \\ 0,8053 & -0,9213 \end{bmatrix}$$

- a) Si stabilisca il numero di funzioni discriminanti da mantenere nell'analisi secondo un criterio puramente descrittivo.
- b) Si calcolino le correlazioni tra le variabili originarie e le funzioni discriminanti e si commentino i risultati ottenuti.
- c) Utilizzando le informazioni opportune si effettui una rappresentazione grafica dell'analisi interpretando adeguatamente.