

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

7 aprile 2006

NB : Commentare sempre i risultati ottenuti.

1. Nella seguente tabella sono riassunti i dati raccolti dal Dipartimento Risorse Umane di una grande azienda relativamente a 300 impiegati rispetto a due caratteristiche, livello di *qualifica* lavorativa e *tipo di copertura assicurativa* posseduta:

<i>Tipo di copertura assicurativa</i> <i>Qualifica</i>	<i>Alto livello</i>	<i>Base</i>	<i>Nessuna</i>	Totale
<i>Impiegato</i>	23	27	50	100
<i>Funzionario</i>	21	29	52	102
<i>A contratto</i>	20	30	48	98
Totale	64	86	150	300

- a) Si descriva il modello log-lineare di distribuzioni di colonna uniformi per la tabella sopra riportata, commentando adeguatamente rispetto al contesto di riferimento.
- b) Si forniscano le stime di massima verosimiglianza delle frequenze attese e dei parametri del modello descritto al punto a).
2. Si consideri la tabella a doppia entrata del precedente esercizio e si valuti la distanza chi-quadrato:
- a) fra i profili del tipo di copertura assicurativa *alto livello* e *nessuna*;
- b) fra il profilo qualifica *impiegato* e il corrispondente profilo medio.
3. Illustrare gli indici sintetici utilizzati nell'analisi delle corrispondenze, fornendo espressione analitica, significato ed impiego.
4. Un archeologo ha analizzato i teschi di 75 scheletri umani ritrovati nel sud del Tibet. In base alle loro caratteristiche anatomiche è stato in grado di classificarli come appartenenti a tre differenti ceppi di razze umane, in particolare: 30 di origine tibetana, 15 di origine mongola e 30 di tipo indiano. Le misurazioni effettuate sui teschi sono state diverse, tra le quali *altezza* (in mm.), *peso* (in gr.), *larghezza zigomi* (in mm.), *altezza del naso* (in mm.). I vettori dei centri dei punteggi discriminanti rispetto alle funzioni discriminanti estraibili sono di seguito riportati:

$$\mathbf{c}_1 = \begin{bmatrix} -0,611 \\ -0,200 \\ +0,711 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} +0,257 \\ -0,746 \\ +0,116 \end{bmatrix}.$$

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

$$\lambda_1 = 0,3741 \quad \lambda_2 = 0,1492,$$

$$\mathbf{A}_{(4 \times 2)} = \begin{bmatrix} 0,110 & -0,140 \\ -0,052 & -0,148 \\ -0,125 & 0,033 \\ 0,153 & 0,275 \end{bmatrix};$$

si stabiliscano:

- a) il numero di funzioni discriminanti da mantenere nell'analisi dal punto di vista inferenziale;
- b) tenendo conto della scelta effettuata al punto precedente, a quale ceppo potrebbe appartenere un ulteriore scheletro ritrovato il cui teschio presenta i seguenti valori delle variabili in esame: 136, 125, 90, 53.