

**Università degli Studi di Milano-Bicocca - Facoltà di Economia**  
**Esame di Analisi dei Dati (modulo B)**  
 30 giugno 2005

*NB : Commentare sempre i risultati ottenuti.*

1. Dimostrare la relazione fra la misura  $X^2$  (di Pizzetti-Pearson) e la distanza chi-quadrato in una tabella di contingenza riferita a due caratteri.
  
2. Una azienda esperta nell'organizzazione di vacanze è interessata alla valutazione della soddisfazione dei proprio clienti ospiti dei villaggi vacanze che l'azienda gestisce. Con tale obiettivo viene selezionato un campione di 150 individui, ospiti dei diversi villaggi nell'ultimo mese, ai quali viene chiesto di valutare - con una scala da 1 a 10 - il servizio offerto dalla struttura del villaggio in relazione a 6 caratteristiche quali la *disponibilità del personale*, la *pulizia delle stanze*, la *qualità del cibo*, il *livello di animazione*, le *offerte sportive* e l'*organizzazione delle escursioni*. Ciascun intervistato esprime la propria soddisfazione globale sulle offerte dei villaggi determinando così la formazione di quattro gruppi:
  - ▶ insoddisfatti;
  - ▶ moderatamente soddisfatti;
  - ▶ pienamente soddisfatti;
  - ▶ entusiasta.

Costruite le matrici  $\mathbf{W}$  e  $\mathbf{B}$  di devianze e codevianze 'nei' e 'fra' i gruppi dei punteggi relativi alle 6 caratteristiche, gli autovalori della matrice  $\mathbf{W}^{-1}\mathbf{B}$  risultano 0,614, 0,311, 0,084. Impiegare tali informazioni per stabilire, da un punto di vista sia descrittivo che inferenziale, il numero di funzioni discriminanti utili per l'indagine avviata dall'azienda.

3. Con l'obiettivo di valutare i servizi in ambito comunale, l'amministrazione centrale ha condotto un'indagine volta ad accertare se il numero di servizi pubblici è ripartito in modo omogeneo tra le 4 zone in cui è suddiviso il comune. Si è analizzato un campione di 200 esercizi pubblici che sono stati classificati secondo la zona di dislocazione e la tipologia di esercizio. I risultati sono riportati nella seguente tabella:

<i>Esercizi pubblici</i>	<i>Zone</i>			
	1	2	3	4
Ristoranti	13	8	9	27
Trattorie	10	7	9	21
Pizzerie	8	6	11	18
Bar-Caffè	9	8	12	24

Fissando  $\alpha = 0,05$ , stabilire quale modello log-lineare gerarchico sia idoneo a descrivere la relazione fra il tipo di esercizio pubblico e la zona di dislocazione nella popolazione di tutti gli esercizi pubblici presenti sul territorio comunale.

4. Si considerino i caratteri  $A$ ,  $B$ , e  $C$ , le cui distribuzioni marginali semplici e doppie di frequenze relative sono rispettivamente:

$A$	$f(a_i)$
$a_1$	0,4
$a_2$	0,6
	<b>1</b>

	$C$		
$B$	$c_1$	$c_2$	
$b_1$	0,1	0,1	0,2
$b_2$	0,6	0,2	0,8
	0,7	0,3	<b>1</b>

- a) Si fornisca la distribuzione congiunta di  $A$ ,  $B$  e  $C$  per la quale il carattere  $A$  sia indipendente dalla distribuzione doppia  $(B, C)$ .
- b) Nella tabella calcolata al punto precedente si mostrino le implicazioni in termini di indipendenza marginale opportuna.