Corsi di Laurea Specialistica in Scienze Ambientali
Corso di Metodi probabilistici e statistici per l'analisi dei dati a.a. 2002-2003
Prova scritta del 24/03/2003

1. (8 punti) (dati cortesemente forniti dal Prof. C. Ferrari)

   Una certa area montana è stata suddivisa in sotto regioni di ampiezze diverse, e le specie caratteristiche della zona sono state classificate. Nella tabella FERRARI.DAT, sono riportate il numero di specie trovate in ogni regione (x_1) e l'area (in mq) della regione (x_2). Si considerino i logaritmi delle variabili. Quindi: 1) Calcolare la matrice di correlazione; 2) Fare un test di normalità bivariata (liv.sign. 5%); 3) Fare un grafico chi-quadro della distanza statistica. Commentare i risultati ottenuti.

2. (12 punti) Viene fatta un'analisi dell'emissione di ossido di carbonio (CO_2) nell'atmosfera. In tabella OSSIDO.DAT sono riportati i dati (nelle appropriate unità di misura) della concentrazione di CO_2 nell'atmosfera, e le emissioni globali di CO_2 di due sorgenti, gas naturali e solidi, nel periodo 1960-1984 (come riportato qui sotto per i primi 3 anni):

Anno	Concentr. CO_2	gas naturale	solidi
1960	316.6	235	1 419
1962	318.2	277	1 358
1964	319.1	328	1 442
$\vdots$	$\vdots$	$\vdots$	$\vdots$

   Fare un'analisi di regressione lineare per la dipendenza di concentrazione di CO_2 rispetto alle due emissioni: determinare una stima del modello e commentare sulla bontà del modello con gli strumenti noti. Determinare una regione di confidenza al 95% per i coefficienti β. Stimare il valore di concentrazione di CO_2 per una emissione di 600 unità da gas naturali e 1600 da solidi (nella loro unità di misura). Valutare infine l'ipotesi che l'emissione dei solidi non influenzi la concentrazione di CO_2.

3. (12 punti) Supponiamo che vengano effettuate $n_1 = 11$ e $n_2 = 12$ osservazioni di due popolazioni π_1 e π_2 aventi distribuzione normale con stessa matrice di covarianza. Le statistiche campionarie sono

   $$\bar{\mathbf{x}}_1 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad S_{com} = \begin{pmatrix} 7.3 & -1.1 \\ -1.1 & 4.8 \end{pmatrix}.$$

   (a) Fare un test T^2 di Hotelling della differenza tra le medie campionarie ($\alpha = 0.10$);

   (b) Costruire la funzione di discriminanza di Fisher;

   (c) Allocare l'osservazione $\mathbf{x}_0^T = [0, 1]$, assumendo uguali costi ed uguali probabilità a priori per le due popolazioni.