

Corso di Laurea Magistrale in Analisi e Gestione dell'Ambiente  
Corso di Analisi Statistica Multivariata a.a. 2010-2011  
Prova scritta del 23/06/2011

1. I dati nella tabella IRIS\_SHAPE si riferiscono alla forma (adimensionale, rapporto tra lunghezza e larghezza dei petali) del fiore delle famiglie *versicolor* e *virginica* di Iris di Fisher. In particolare, sono riportate la forma del sepalò ( $x_1$ ) e quella del petalo ( $x_2$ ); la terza colonna rappresenta la famiglia. Si vuole valutare se le due popolazioni di forme sono effettivamente distinte.
  - (a) fare un test di normalità univariata e bivariata delle due popolazioni (liv. sign.  $\alpha = 0.01$ );
  - (b) Fare un confronto delle medie delle due popolazioni individuando la regione di confidenza per la differenza delle medie (liv. sign.  $\alpha = 0.01$ );
  - (c) Dopo aver effettuato un'analisi di discriminanza delle due popolazioni, allocare il nuovo campione:  $\mathbf{x}_0 = [2, 2.8]$ . Qual'è il valore di APER? Commentare.
2. I dati della Tabella **Columbus** (in DASL and MORE DATA) si riferiscono a dati di criminalità nella città di Columbus, Ohio.
  - (a) Mediante un'analisi di regressione lineare, valutare la dipendenza lineare della risposta CRIME, dalle variabili predittrici AREA, HOVAL, INC, OPEN, PLUMB. In particolare: determinare una stima del modello e valutare la bontà del modello con gli strumenti noti.
  - (b) Stimare il numero di crimini per i seguenti valori delle variabili predittrici: [0.50 24 10 0.42 0.70].
  - (c) Valutare l'indice di correlazione spaziale di Moran per il residuo del modello, usando X, Y per le coordinate spaziali (liv.sign. 5%, 1%).
  - (d) Valutare l'ipotesi che la variabile OPEN non influenzi il numero di crimini (livello di sign. 5%).