

Corso di Laurea Specialistica in Scienze Ambientali
Corso di Metodi probabilistici e statistici per l'analisi dei dati a.a. 2006-2007
Prova scritta del 12/03/2007

1. I dati nella tabella IRIS.SHAPE si riferiscono alla forma (adimensionale, rapporto tra lunghezza e larghezza dei petali) 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 diagramma di dispersione e commentare.
 - (b) fare un test di normalità univariata e bivariata delle due popolazioni (liv. significatività $\alpha = 0.01$;
 - (c) Fare un confronto delle medie delle due popolazioni, valutando sia la regione di confidenza che gli intervalli simultanei di confidenza per la differenza delle medie, considerando un livello di significatività $\alpha = 0.01$;
 - (d) Dopo aver effettuato un'analisi di discriminanza delle due popolazioni, allocare il nuovo campione: $\mathbf{x} = [2, 2.8]$. Qual'è il valore di APER? Commentare.
2. I dati della Tabella GAS.DAT si riferiscono alle osservazioni di una azienda di elettricità per la pianificazione della propria produzione. Le variabili delle 63 osservazioni giornaliere rappresentano:

Y Gas Naturale distribuito ai clienti

X_1 Misura legata alla temperatura giornaliera: DHD = 65 deg - temp.media giorn.

X_2 Velocità del vento (media giornaliera)

X_3 weekend/non-weekend

Fare un'analisi di regressione lineare per un modello di dipendenza di Gas Naturale distribuito ed i parametri giornalieri considerati. In particolare: determinare una stima del modello e commentare sulla bontà del modello con gli strumenti noti. Stimare il fabbisogno di Gas per valori delle variabili predittrici di [38, 15, 0].

Valutare infine l'ipotesi che la variabile weekend non influenzi il fabbisogno di Gas (livello di sign. $\alpha = 0.05$).