

Corsi di Laurea Specialistica in Scienze Ambientali
Corso di Metodi probabilistici e statistici per l'analisi dei dati a.a. 2003-2004
Prova scritta del 15/03/2004

1. In tabella METEO.DAT sono riportate le osservazioni meteorologiche relative ad un periodo di 11 anni (dal 1920-21 al 1930-31) in una zona agricola inglese. Le variabili rappresentano
 - x_1 Pioggia caduta in Novembre e Dicembre (mm.)
 - x_2 Temperatura media di Luglio (gradi c.)
 - x_3 Pioggia caduta in Luglio (mm.)
 - x_4 Radiazione in Luglio (ml.)
 - (a) (4 punti) Determinare la matrice di correlazione e commentare i risultati.
 - (b) (12 punti) Determinare le componenti principali con la matrice di correlazione e con quella di covarianza. Studiare la varianza spiegata dalle componenti nei due casi. Confrontare i risultati.

2. Il gene *Chaetocnema* contiene due tipi di specie, Ch.concinna e Ch. heikertingeri, che sono state confuse tra loro per molto tempo. Il tabella GENI.DAT sono riportate alcune caratteristiche delle due popolazioni, x_1, x_2 per Ch. concinna e x_3, x_4 per Ch. heikertingeri, relative a due dimensioni significative (larghezza della prima sezione e della seconda sezione del gene).
 - (a) (3 punti) Per la popolazione Ch. concinna, fare un test di normalità bivariata.
 - (b) (5 punti) Fare un test T^2 di Hotelling della differenza tra le medie campionarie delle due popolazioni ($\alpha = 0.10$), sotto le usuali ipotesi di normalità e di uguali matrici di covarianza delle due popolazioni.
 - (c) (8 punti) Determinare la funzione di discriminanza di Fisher. Allocare l'osservazione $\mathbf{x}_0^T = [190, 125]$, assumendo uguali costi ed uguali probabilità a priori per le due popolazioni. Riportare le osservazioni su un grafico.