

Corsi di Laurea Specialistica in Scienze Ambientali
Corso di Metodi probabilistici e statistici per l'analisi dei dati a.a. 2006-2007
Prova del 24/07/2007

1. Si considerino i dati relativi ai records femminili di atletica di varie nazioni (`t1-9-x.dat` e foglio allegato).
 - (a) Standardizzare le variabili
 - (b) Fare un'analisi di clustering per le discipline e per le nazioni, con gli strumenti noti. Commentare i risultati.
 - (c) Calcolare la matrice di correlazione e commentare.
 - (d) Supponendo una distribuzione normale della popolazione, fare un test di fattibilità di un'analisi fattoriale per tali variabili (liv. 5%).
 - (e) Fare un'analisi fattoriale ed interpretare i fattori determinati.

2. Si considerino i dati di insetti *Leptoconops* (`T6-15.DAT`) che si riferiscono alle due famiglie morfologicamente simili *L.torrens* ($n_1 = 35$) e *L. carteri* ($n_2 = 35$). In totale ci sono 7 variabili.
 - (i) Fare un confronto delle medie delle due popolazioni, valutando gli intervalli simultanei di confidenza per la differenza delle medie, considerando un livello di significatività $\alpha = 0.05, 0.01$. Commentare i risultati
 - (ii) Si considerino la prima variabile (Wing length) e la terza variabile (Third palm length). Dopo aver effettuato un'analisi di discriminanza delle due popolazioni, allocare il nuovo campione: $\mathbf{x}_0 = [92, 37]$. Qual'è il valore di APER? Commentare.
 - (iii) Discutere i risultati di un'analisi di discriminanza utilizzando le prime due variabili (Wing length e Wing width). I risultati sono inattesi?