

Corsi di Laurea Specialistica in Scienze Ambientali  
Corso di Metodi probabilistici e statistici per l'analisi dei dati a.a. 2005-2006  
Prova scritta del 23/03/2006 ore 9.00

1. Si considerino i dati relativi ai records (maschili) in atletica di varie nazioni (`t8-6-x.dat` e foglio allegato).
  - (a) Fare un'analisi di clustering per le variabili con gli strumenti noti. Commentare i risultati.
  - (b) Supponendo una distribuzione normale della popolazione, fare un test di fattibilità di un'analisi fattoriale per tali variabili. L'ipotesi di normalità è ragionevole?
  - (c) Fare un'analisi fattoriale ed interpretare i fattori determinati.
  
2. Si considerino le  $n = 50$  osservazioni relative alla famiglia di *iris versicolor* del data set di R. Fisher (file IRIS1). Le due variabili rappresentano la lunghezza del sepalo ( $x_1$ ) e del petalo ( $x_2$ ).
  - (a) Determinare la regione (ellisse) di confidenza al 95% per la *media* delle due variabili. Determinare inoltre gli intervalli simultanei ( $T^2$ ) di confidenza al 95% e 90% per la media delle due variabili. Riportare i grafici in un diagramma di dispersione.
  - (b) Si consideri quindi anche il campione di  $n = 49$  individui della famiglia di *iris virginica* (IRIS2). Confrontare le matrici di covarianza. Fare un test di uguaglianza delle medie, con un livello di significatività del 5% e dell' 1%.
  - (c) Supponendo uguali costi e probabilità, allocare la nuova osservazione  $\mathbf{x} = [6.4, 4.6]$  in una delle due famiglie, mediante la retta discriminante di Fisher. Qual'è il coefficiente di APER?