

Corsi di Laurea Magistrale in Scienze Ambientali
Corso di Analisi Statistica Multivariata a.a. 2008-2009
Prova scritta del 08/07/2009 ore 8.30

1. Si considerino i dati di *air-pollution* (T1-5.DAT), dove x_1 : vento, x_2 : radiazione solare, x_3 : CO, x_4 : NO, x_5 : NO₂, x_6 : O₃, x_7 : HC.
 - (a) Fare un'analisi di clustering delle variabili mediante metodi di *linkage* e Multidimensional Scaling, analizzando diverse possibili distanze. Commentare i risultati.
 - (b) Supponendo una distribuzione normale della popolazione, fare un test di fattibilità di un'analisi fattoriale per tali variabili.
 - (c) Fare un'analisi fattoriale: valutare il numero minimo di carichi necessari per una buona rappresentazione della variabilità dei dati, ed interpretare i fattori determinati.

2. Si considerino le prime $n = 36$ osservazioni relative alla documentazione aziendale dei costi associati all'uso di camion a "gasoline" (Tab. T6-10.DAT). Le tre variabili rappresentano il consumo di benzina (x_1), i costi per riparazioni (x_2) e il capitale investito (x_3) (il tutto per miglio). Considerare per tutta l'analisi che segue **solo le prime due variabili**.
 - (a) Mediante un diagramma di dispersione tra le variabili (x_1, x_2) identificare eventuali outliers ed eliminarli.
 - (b) Effettuare un test di normalità **univariata** per entrambe le variabili.
 - (c) Determinare la regione (ellisse) di confidenza al 95% e 99% per la *media* delle due variabili. Determinare inoltre gli intervalli simultanei (T^2) di confidenza al 95% e 90% per la media delle stesse due variabili. Riportare tutti i grafici in un diagramma di dispersione.
 - (d) Si consideri quindi anche il campione di $n = 23$ osservazioni per camion *diesel*. Confrontare la matrice di covarianza con quella del precedente campione. Fare un test di uguaglianza delle medie (liv. sign. 5% e 1%), motivando l'approccio usato.
 - (e) Supponendo uguali costi e probabilità, allocare la nuova osservazione $\mathbf{x} = [12.4, 9.2]$ mediante una procedura di discriminanza. Qual'è il coefficiente di APER?