

Analisi Statistica Multivariata. A.A. 2012-2013
Progetto n.1. Gruppo VI. Consegna: 6/6/2013.
I dati per i problemi sono reperibili sul sito del corso

Problema 1.

Si considerino i dati **Pollution-altro set** (in *DASL and More Data*).

1. Calcolare la matrice di correlazione e commentare.
2. Fare un'analisi di clustering usando single e complete linkage, considerando almeno tre distanze, commentando sulle scelte fatte. Interpretare i risultati alla luce del problema.
3. Valutare i raggruppamenti con anche un metodo di Multidimensional scaling, e confrontare i risultati.

Problema 2.

Si considerino i dati di *Mineral Content*, nella Tabella 6.16 (T6-16.DAT in JW). Si considerino le due famiglie "osso dominante" e "osso non dominante" (le variabili sono : radius, humerus, ulna).

1. Per la matrice di osservazioni relativa ad ogni popolazione considerata, fare uno studio della normalità univariata, bivariata e multivariata.
Verificare la presenza di outliers e, nel caso, eliminarli dall'analisi successiva giustificando la scelta. Trasformare eventualmente le variabili, o alcune di esse, per migliorare la normalità dei dati.
2. Per le osservazioni provenienti dalle popolazioni, valutare l'ipotesi di uguale media delle popolazioni, con livello di significatività $\alpha = 0.05$ e $\alpha = 0.01$. In caso di rifiuto, raffinare l'analisi facendo un confronto di coppie di variabili, individuando anche la direzione di massima differenza. Commentare i risultati e riportare i risultati su un grafico.
3. Nel confronto delle medie, determinare intervalli simultanei di confidenza (95%) e commentare su eventuali differenze tra i risultati ottenuti rispetto al test del quesito precedente. Riportare i risultati su un grafico.
4. Impostare un test di discriminanza per allocare la nuova osservazione

$$\mathbf{x} = [0.81, 2.0, 0.78]$$