

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

Problema 1.

Si considerino i dati **Jobs** (in *DASL and More Data*). Sia per le variabili che per le osservazioni, fare quanto segue:

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 *Hook-billed Kites*, nella Tabella 5-12 (T5-12.DAT in JW).

1. Fare uno studio della normalità univariata, bivariata.
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. Determinare e rappresentare su un grafico la regione di confidenza per la media (al 99% e 95%).
3. Mediante un test con la T^2 di Hotelling, valutare se il vettore

$$\mu_0 = [208.5, 270]$$

potrebbe essere un buon candidato per la media della popolazione (liv.sign.1% e 5%). (riportare sia il risultato del test che la valutazione grafica)

4. Determinare intervalli simultanei di confidenza (95% e 99%) per le singole variabili. Riportare i risultati sul grafico del punto 2.