

Analisi Statistica Multivariata. A.A. 2011-2012
Progetto n.1. Gruppo VI. Consegna: 30/5/2012.
I dati per i problemi sono reperibili sul sito del corso

Problema 1.

Si considerino i dati **Large - Piante** (in *DASL and More Data*), riferiti alla distribuzione di alcuni tipi di piante in una regione montuosa suddivisa in settori.

Dopo aver valutato se studiare gli oggetti e/o le variabili, fare un'analisi di clustering usando single e complete linkage, eventualmente considerando varie distanze. Interpretare i risultati indicando le eventuali eccezioni.

Valutare i raggruppamenti con anche un metodo di Multidimensional scaling, e confrontare i risultati.

Problema 2.

Si considerino i dati **Jellyfish** (in *DASL and More Data*).

1. Per la matrice di osservazioni relativa ad ogni popolazione considerata, fare uno studio della normalità univariata e bivariata. Trasformare eventualmente le variabili, o alcune di esse, per migliorare la normalità dei dati. Controllare se ci sono osservazioni anomale o outliers.
2. Determinare la regione di confidenza (99% e 95%) per la media, relativamente ad entrambe le famiglie. Riportare tutto sullo stesso grafico e commentare.
3. Per le osservazioni provenienti dalle diverse popolazioni, valutare l'ipotesi di uguale media delle popolazioni, con livello di significatività $\alpha = 0.05$ e $\alpha = 0.01$. Commentare i risultati. In caso di rifiuto, valutare quale delle variabili è più responsabile del rifiuto, facendo il confronto di medie sulle singole variabili.
4. Determinare intervalli simultanei di confidenza (95%) e commentare su eventuali differenze tra i risultati ottenuti rispetto al test del quesito precedente.
5. Impostare un test di discriminanza per allocare la nuova osservazione

$$\mathbf{x} = [12, 15].$$