

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

Problema 1.

Si considerino i dati **elnino** (in *DASL and More Data*). 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 **Pottery** (in *DASL and More Data*) per le famiglie corrispondenti ai siti.

1. Per la matrice di osservazioni relativa ad ogni popolazione considerata, fare uno studio della normalità univariata e multivariata (test chi²). Verificare se siano presenti outliers: in caso affermativo, 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 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.
3. Determinare intervalli simultanei di confidenza (95%) e commentare su eventuali differenze tra i risultati ottenuti rispetto al test del quesito precedente.
4. Sfruttare il metodo di discriminanza principale di Fisher per una valutazione grafica delle famiglie e confrontare con i precedenti risultati.
5. Impostare un test di discriminanza per allocare la nuova osservazione

$$\mathbf{x} = [17.1, 1.84, 0.60, 0.13, 0.05, *]$$