

Analisi Statistica Multivariata. A.A. 2011-2012

Progetto n.1. Gruppo X. 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 **FishCatch** (in *DASL and More Data*) per le famiglie di Lahna (1) e Ahven (7). Considerare solo le variabili 3 (Weight), 4 (Length1), 5 (Length2) e 6 (Length3) nel testo. Gestire eventuali valori mancanti ("missing data").

1. Per la matrice di osservazioni relativa ad ogni popolazione considerata, fare uno studio della normalità univariata e multivariata (test chi2). Verificare la presenza di outliers ed eliminarli dall'analisi successiva giustificando la scelta. Trasformare eventualmente le variabili, o alcune di esse, per migliorare la normalità dei dati.
2. Determinare intervalli simultanei di confidenza (99% e 95%) per la media, relativamente alla famiglia Ahven.
3. Per le osservazioni provenienti dalle due 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 sia 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. Sfruttare il metodo di discriminanza principale di Fisher per una valutazione grafica delle famiglie e confrontare con i precedenti risultati.
6. Impostare un test di discriminanza per allocare la nuova osservazione (vengono date tutte le variabili)

$$\mathbf{x} = [*, 800.0, 33.7, 36.4, 39.6, 29.7, 16.6, 0].$$