

Analisi Statistica Multivariata. A.A. 2011-2012

Progetto n.1. Gruppo IX. Consegna: 30/5/2012.

I dati per i problemi sono reperibili sul sito del corso

Problema 1.

Si considerino i dati **Sleep of Animals** (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. Fare attenzione ed eventualmente gestire possibili *outliers*.

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

Problema 2.

Si considerino i dati di **Biting Flies** (in *T6-15.DAT in JW*) relative a due tipi di mosche (*Leptoconops carteri* e *Leptoconops torrens*) che per lungo tempo erano state considerate la stessa famiglia. Le variabili si riferiscono a (dalla tabella): Wing length, Wing width, Third palp length, Third palp width, Fourth palp length, Length of antennal segment 12, Length of antennal segment 13. (si veda anche il libro di testo, p.347).

1. Per la matrice di osservazioni relativa ad ogni popolazione considerata, fare uno studio della normalità univariata e multivariata (test χ^2). 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

$$\mathbf{x} = [100, 48, 38, 14, 32, 10, 9]$$