

Analisi Statistica Multivariata
 Esercizio di Laboratorio dell' 11/4/2012
<http://www.dm.unibo.it/~simoncin/DatiII.html>

- Data la matrice di dati della Tabella 11.5 (T11-5.DAT in JW) si considerino le tre famiglie di IRIS (setosa, versicolour, virginica) con le seguenti variabili:

Lung. Sepalo	Larg. Sepalo	Lung. Petalò	Larg. Petalò	Famiglia
5.1	3.5	1.4	0.2	1
4.9	3.0	1.4	0.2	1
⋮	⋮	⋮	⋮	⋮

1. Per le osservazioni di ogni popolazione, determinare i vettori media (funzione `mean` di Matlab), le matrici di covarianza (funzione `cov` di Matlab) e le matrici di correlazione (funzione `corrcoef` di Matlab). Commentare i valori delle 3 matrici di correlazione.
 2. Fare il diagramma di dispersione (`plot`) per le variabili (x_1, x_3) (lunghezza sepalò e lunghezza petalò) delle famiglie *setosa* e *virginica* (i due diagrammi sullo stesso grafico, con valori degli assi e legenda). Commentare i risultati in relazione ai corrispondenti coefficienti di correlazione.
- Sono state individuate 6 zone con muschio, ed è stata segnalata la presenza di 4 specie di muschi (i dati sono nel sito):

Specie \ zona	1	2	3	4	5	6
Oxycoccus microcarpon	1	1	0	1	0	1
Sphagnum magellanicum	0	1	1	0	1	0
Sphagnum fuscum	1	1	0	1	0	0
Smilacina trifoliata	0	1	1	0	1	1

Calcolare i coefficienti di dissimilarità (funzione `dissim.m` nel sito) mediante le distanze 'Jaccard', 'Euclidean', 'Hamming'. Fare l'analisi sia per le zone che per le specie. Commentare i risultati.