

Corso di Laurea Magistrale in Analisi e Gestione dell'Ambiente
Corso di Analisi Statistica Multivariata a.a. 2012-2013
Prova scritta del 04/07/2013

1. I dati nella tabella **turtles** si riferiscono alla misure di larghezza e altezza dei carapaci di due campioni di tartarughe (femmine e maschi).
 - (a) Fare un diagramma di dispersione dei due campioni e commentare.
 - (b) Fare un test di normalità univariata e bivariata dei due campioni (liv. sign. $\alpha = 0.05$) (non effettuare trasformazioni);
 - (c) Determinare la regione di confidenza (95% e 99%) per la media di ognuna delle due popolazioni. Riportare tutti i grafici sulla stessa figura;
 - (d) Fare un confronto delle medie delle due popolazioni, valutando sia la regione di confidenza che gli intervalli simultanei di confidenza per la differenza delle medie, considerando un livello di significatività $\alpha = 0.05$;
 - (e) Dopo aver effettuato un'analisi di discriminanza delle due popolazioni, allocare il nuovo campione: $\mathbf{x}_0 = [90, 45]$. Qual'è il valore di APER? Commentare.

2. I dati nella tabella T8-5.DAT (JW) si riferiscono ad informazioni su diversi distretti di una contea americana. Le variabili riportate sono
 - x_1 : num.abitanti
 - x_2 : scolarizzazione (numero di anni)
 - x_3 : numero di lavoratori (in migliaia)
 - x_4 : numero di lavoratori nel servizio sanitario (in centinaia)
 - x_5 : valore mediano delle abitazioni (in 10,000 USD)
 - (a) Calcolare la matrice di correlazione e commentare.
 - (b) Effettuare un'analisi della componenti principali: analizzare la varianza spiegata dalle componenti principali ed il coefficiente di correlazione tra le nuove e vecchie variabili. Valutare l'eventuale presenza di osservazioni anomale. Valutare il numero di componenti principali sufficienti per una buona descrizione delle osservazioni.
 - (c) Assumendo che le variabili originarie provenissero da una distribuzione normale, possiamo accettare l'ipotesi che ognuna delle prime due componenti principali provenga da una distribuzione normale?
 - (d) Interpretare i risultati, possibilmente individuando una caratteristica di ognuna delle componenti principali tenute. Riportare le osservazioni su di un grafico delle componenti principali tenute, e commentare;