

Corsi di Laurea Magistrale in Scienze Ambientali
Corso di Analisi Statistica Multivariata a.a. 2008-2009
Prova scritta del 27/07/2009 ore 8.30

1. La tabella T9-12.DAT contiene i risultati di una valutazione dell'attività e di un test attitudinale su un gruppo di $n = 50$ venditori di una azienda. Le variabili si riferiscono a: x_1 : Crescita nelle vendite; x_2 : Capacità di vendita; x_3 : Vendite a nuovi acquirenti; x_4 : Creatività; x_5 : capacità meccaniche; x_6 : capacità di astrazione; x_7 : abilità matematiche. Effettuare una trasformazione $\tilde{x} \leftarrow \sqrt[4]{x}$, Determinare la matrice di correlazione e commentare i valori trovati. Usare le variabili trasformate nel seguito dell'esercizio.
 - (a) Fare un'analisi di clustering delle variabili mediante metodi di *linkage* e Multidimensional Scaling, analizzando diverse possibili distanze. Commentare i risultati.
 - (b) Supponendo una distribuzione normale della popolazione, fare un test di fattibilità di un'analisi fattoriale per tali variabili.
 - (c) Fare un'analisi fattoriale: valutare il numero minimo di carichi necessari per una buona rappresentazione della variabilità dei dati, ed interpretare i fattori determinati.

2. I dati nella tabella FISH_F si riferiscono al diametro (in centesimi di pollice) degli anelli di un campione di salmone femmina dopo un anno in acqua dolce (x_1) e dopo un anno in oceano (x_2), per due popolazioni di salmone: Canadese (popolazione 1) e dell'Alaska (popolazione 2). Si vuole valutare se le due popolazioni sono effettivamente distinte.
 - a) Per ogni variabile, effettuare un test di normalità *univariata* delle due popolazioni;
 - b) Prima del confronto delle due popolazioni, determinare **per ogni popolazione** la regione (ellisse) di confidenza al 95% e 99% per la *media* delle variabili. Riportare tutti i grafici in un unico diagramma di dispersione insieme ai dati.
 - c) Fare un test di 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 $\alpha = 0.05$;
 - d) Dopo aver effettuato un'analisi di discriminanza delle due popolazioni, allocare il nuovo campione: $\mathbf{x} = [80, 400]$. Qual'è il coefficiente di APER?