

Corsi di Laurea Specialistica in Scienze Ambientali
Corso di Metodi probabilistici e statistici per l'analisi dei dati a.a. 2003-2004
Prova scritta del 06/07/2004

1. (12 punti) I valori in tabella PRECAPITA.DAT si riferiscono a dati relativi ai salari medi (in USD) dell'anno 1960 (prima colonna), di alcuni paesi dell'occidente (es. Canada, Svezia, Francia, ecc.). Nella Tabella sono anche presenti per ogni paese, le percentuali rispettivamente di forza lavoro impiegata in agricoltura, nell'industria e nel terziario.

Mediante un Modello di regressione lineare, studiare la dipendenza dei salari medi dall'impiego di forza lavoro nei vari settori. In particolare

- (a) Determinare la matrice di correlazione delle variabili predittrici. Commentare.
 - (b) Determinare una stima corretta per i parametri del modello.
 - (c) Valutare la bontà del modello mediante tutti gli strumenti noti (coeff. determinazione, studio grafico del residuo).
 - (d) Fare un test di significatività ($\alpha = 0.5$) sulla variabile predittrice x_1 .
2. La tabella POTTERY.DAT contiene un campione di cocci di era romanica rinvenuti in alcuni siti in Gran Bretagna. Le variabili rappresentano la percentuale di ossido di vari metalli, in particolare x_1 : ossido di alluminio, x_2 : ossido di ferro, x_3 : ossido di magnesio, x_4 : ossido di calcio e x_5 : ossido di sodio.
- (a) (8 punti) Fare un test χ^2 di normalità multivariata. Commentare i risultati. Fare un test χ^2 escludendo la prima variabile. Commentare i risultati.
 - (b) (12 punti) Effettuare un'analisi fattoriale col metodo dei fattori principali, descrivendo i pesi stimati L_1, \dots, L_m dei fattori, e le varianze specifiche. Per $m = 2$, eventualmente individuare una rotazione dei carichi che possa aiutare l'interpretazione dei risultati. Stimare i fattori per $m = 2$ relativamente alla prima osservazione.