02424 Week 4

The main purpose of the exercises this week is to illustrate aspects in relation to formulation and testing of general linear models. Important aspects are formulation of the design matrix, least squares estimation, testing for parameters, and model structure.

Exercise 1

Effect of measuring wind speed at a new location

In an airport near Copenhagen the, wind speed is measured every hour using a cup-anemometer. Suppose that now ($t = 8$) the following observations of the wind speed are given [m/s]:

$$4.4, 3.4, 3.3, 2.5, 7.3, 4.9, 4.8, 4.4$$

Before the fifth measurement was performed, the anemometer was moved from the normal measuring location (2m above ground level) to the roof of a building.

Question 1

Formulate a suitable model describing the variations in the observed wind speed.

Question 2

Estimate the parameters of the model and especially assess (model) the difference in wind speed at the two measurement locations.

Question 2

Predict the wind speed at the old measuring location at the next hour ($t = 9$).

Exercise 2

Air Pollution (SO$_2$) in Cities

It is well-known that air pollution has serious consequences for the health of people and for the life time of materials; in particular those used in buildings. Sulfur dioxide (SO2) is often considered as one of the most important indicators of air quality. SO2 is primarily formed from the combustion of sulphur in fuels, and in addition to the known degeneration of materials, SO2 can in general seriously effect the health of people and have a negative impact on the environment. Air pollution in a city is related both both the size and the city and to climate variables. Modelling is important for a better understanding and prevention of air pollution.

In this exercise we shall try to obtain a better understanding of the relation between the concentration of sulphur dioxide in US cities and its relation to both human ecology and four climate variables.

The file 'so2.txt' contains the annual mean concentration of SO2 (in micrograms per cubic metre) for 41 US cities. The data generally are constructed as the means over the three years 1969–71.

The variables in the file 'so2.csv' are

- Column 1 (Pollution): Mean SO2 concentration

- Column 2 (Temp): Average annual temperature in degrees Fahrenheit

- Column 3 (Industry): Number of manufacturing enterprises employing 20 or more workers

- Column 4 (Population): Population size (1970 values) in thousands

- Column 5 (Wind): Average annual wind speed in miles per hour

- Column 6 (Rain): Average annual precipitation in inches

- Column 7 (Wet.days): Average number of days with precipitation per year

Question 1

Use scatter-plots and correlations to examine the relationships among the individual variables.

Question 2

Establish a model describing the SO2 pollution in US cities. Discuss briefly the estimated influence of the various parameters.

Question 3

In reality the air pollution in a city might be influenced from the air pollution in nearby cities. Describe a method for taking this correlation between nearby cities into account.