MAS programmes - Statistical Analysis (Autumn Term)

Exercises 3

1. Recall the oil example introduced in Lecture 3.

   (a) The *chosen* model uses the explanatory variables `distil` and `endpoint` to predict the response `spirit`.

   Carry out an exploratory analysis of the data, examining the relationships between the response variable `spirit` and the four explanatory variables `gravity`, `pressure`, `distil` and `endpoint`. On the basis of this analysis, which variables might you have expected to see as explanatory variables in a good linear regression model. Does the *chosen* model appear reasonable?

   You may find the following functions useful.

   ```
   > pairs(oil)
   ```

   ```
   > cor(oil)
   ```

   (b) We would like to obtain a prediction for the percentage yield of petroleum spirit (`spirit`) based on a linear model containing the two explanatory variables `endpoint` and `distil` which are assumed to take the values 400 and 200 respectively.

   Using the argument `interval = c("confidence")` and `interval = c("prediction")` in `predict()` calculate the corresponding 95% confidence and predictive interval.

2. Recall the sugar example of Exercises 2, where you were asked to investigate the relationship between sugar consumption `consump` (measured in pounds per capita) and its price `price` (in cents per pound) over the period 1875 to 1929.

   (a) Produce an ANOVA table for the hypothesis that there is no regression relationship between `lconsump` and `price`, and carry out the test.

   The result of this test is reported in two different places on the R summary output from the fitted linear model. Can you identify both these instances?

   (b) Use the fitted model to calculate a 95% *prediction interval* for the predicted consumption of sugar per capita in which the price of sugar is 6 cents per pound (that you calculated '*by-hand*' in Exercises 2).