MAST7866 Linear Regression

Computing Session 4

For this session, we will use three data sets: `nyc.csv`, `beams.txt`, and `defects.txt`. The data is available from the module Moodle page.

1 New York food price data

The New York food price data is stored in the csv (comma separated) file `nyc.csv`. Assuming that you have saved the file in your working directory, it can be loaded using the command

```
nyc <- read.csv("nyc.csv", header=T)
```

(a) Reproduce the results from the lecture.

(b) Calculate predictions for a restaurant with Food score of 25, Decor score of 21 and Service of 22 using

  (i) the model with Food, Decor and Service,

  (ii) the model with Food and Decor only,

  (iii) the model with Food and Service only,

  (iv) the model with Food only.

2 Strength of beams data

The file `beams.txt` contains data on ten wood beams. The variables are: the strength of the wood beams, measurements of specific gravity and moisture content. The units of measurement are unknown. We are interested in the effect of the specific gravity and the moisure content on the strength of a beam. You should load the data into R as the dataframe beams using the `read.table` command.

(a) Draw scatter plots of all variables using the command pairs(beams). What do the plots indicate about the relationships between the variables?

(b) Write down the estimated regression equation and the R^2 for the following regression models. How would you interpret the estimated parameters in each case?

  (i) Strength as response and specific gravity as an explanatory variable.

  (ii) Strength as response and moisture as an explanatory variable.

  (iii) Strength as response and specific gravity and moisture as explanatory variables.

  (iv) How do the values of R^2 compare for the three models in parts (i), (ii) and (iii)?

(c) Check the fit of each model by plotting residuals using the plot function.

(d) Calculate the prediction interval when moisture is 10.1 and specific gravity is 0.5 using

  (i) the model with moisture only,

  (ii) the model with specific gravity only,

  (iii) the model with both moisture and specific gravity.

3 Defective rates data

Siegel (1997, pp. 509–510) claims: "Everybody seems to disagree about just why so many parts have to be fixed or thrown away after they are produced. Some say that it's the standard deviation of the temperature of the production process, which needs to be minimised. Others claim it is clearly the density of the product, and that the problems would disappear if the density is increased. Then there is Ole, who has been warning everyone forever to take care not to push the equipment beyond its limits. This problem would be easiest to fix, simply by slowing down the production rate; however, this would increase some costs. The table below gives the average number of defects per 1,000 parts produced (denoted by Defective) along with values of the other variables described above for 30 independent production runs."

The data are stored in the file `defects.txt`. There is one response variable, Defective, and three explanatory variables: Temperature, Density and Rate. You should load the data into R and fit a linear regression model with sqrt(Defective) as the response variable.

(a) Look at plots of the residuals. Do these indicate that the linear regression model is suitable?

(b) Use the summary function to look at the results of the regression. What can you conclude about the effects of Temperature, Density and Rate?

(c) Use the t-test for the effect of Rate given that Temperature and Density have already been included in the model. What is the result of the test?

(d) Returning to the text at the start of this section:

  (i) Is Ole's theory correct?

  (ii) Does increasing the density decrease the defective rate?