Statistical Modeling I
Practical in R

Practical in R

In this practical, we will work with the Liver dataset (liver.csv). We will look at two different models and their analysis.

A hospital surgical unit was interested in predicting survival in patients undergoing a particular type of liver operation. A random selection of 54 patients was available for analysis. From each patient record, the following information was extracted from the pre-operation evaluation and are reported in the liver.csv file in the following order:

- X_3: blood clotting score

- X_2: prognostic index

- X_1: enzyme function test score

- $log_{10}y$: the base 10 logarithm transformation of the response variable (Y), which is the number of weeks the patients survived after the operation.

1. Read these variables into R calling them `x1, x2, x3, ly`.

2. Fit Model 1: $\log_{10} Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$, where $\varepsilon_i \underset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Obtain the fitted model using summary. Save the fitted values and standardised residuals.

3. Assess the assumptions of normality and constant variance of the random errors by looking at suitable plots.

4. Obtain the scatter plot of the standardised residuals versus explanatory variable X_2. Is there a possible relationship between the residuals and X_2?

5. Fit Model 2: $\log_{10} Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$, where $\varepsilon_i \underset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, Obtain the fitted model and save values as before.

6. Assess the assumptions of normality and constant variance of the random errors.

7. Obtain the scatter plot of the standardised residuals from the model including X_1 and X_2 (as in point 4) versus explanatory variable X_3. Is there a possible relationship between the residuals and X_3?