Statistical Modeling I
Practical in R

Practical in R

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

The data are obtained in a production process of oxidizing ammonia. The variables of interest are:

- Y: the stack loss, which is the percentage of the ingoing ammonia that escapes unabsorbed;

- X_1: the airflow

- X_2: the cooling water inlet temperature in degrees C;

- X_3: the acid concentration in percent.

1. Fit Model 1: $Y_i = \beta_0 + \beta_1 x_{1,i} + \beta x_{2,i} + \beta_3 x_{3,i} + \varepsilon_i$ with $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

2. Check if there are any apparent problems with the residuals;

3. Test the hypothesis regarding the overall regression by using the F test

4. Test the hypothesis regarding the parameters β_j for $j = 0, 1, 2, 3$ by using the t tests

5. Fit Model 2: $Y_i = \beta_0 + \beta_1 x_{1,i} + \beta x_{2,i} + \varepsilon_i$ with $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$

6. Check if there are any apparent problems with the residuals;

7. Test the hypothesis regarding the overall regression by using the F test

8. Test the hypothesis regarding the parameters β_j for $j = 0, 1, 2$ by using the t tests

9. Which is the best model between Model 1 and Model 2 and why?