Ignacio Cascos
Simulation in Probability and Statistics
BSc Applied Maths and Computing
Universidad Carlos III de Madrid

Bootstrap project

File `data.zip` contains 37 data files, each of them named `data_NIU.csv`, where `NIU` is the NIU of one of the students in the group. Form sixteen groups of 2 students and one group of 3 students and read the data file with the NIU of one of those students (any of them) as `read.csv("data_NIU.csv",header=TRUE)`.

Dataset `data_NIU.csv` contains data on variables `y`, `x1`, `x2`, `x3` and the goal is to build a linear regression model to predict `y` in terms of some of the other three variables. The dataset contains outliers and after building any OLS model it can be checked that the error residuals are not normally distributed. For these reasons, it makes sense to build a robust linear regression model (`rlm`) and use the bootstrap to study the significance of the regressors.

1. Build a robust linear regression model with the three covariates and use (95%) bootstrap confidence intervals on the regressors' coefficients to study their significance.

2. Use backward elimination to select the relevant covariates and provide the chosen regression model. Start with the three covariates and delete the ones that do not contribute significantly to the model (as suggested by the bootstrap confindence intervals) until all remaining covariates contribute significantly to the regression model.

3. Provide 95% confidence intervals on the regression coefficients of the chosen model in the previous part.

4. Build a 95% confidence interval on the mean response when $(\mathtt{x1}, \mathtt{x2}, \mathtt{x3}) = (14, 14, 14)$.