Computational Statistics II
Assignment 1: Bootstrap for linear regression

Alessia Pini

The R dataset `mtcars` contains data extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). You can access the data with the command `data(mtcars)`. We are interested in exploring the relationship between a set of variables and miles per gallon (MPG) as outcome.

1. Give a theoretical description of the Bootstrap for regression, specifying the model and assumptions. Describe the two methods of bootstapping residuals and pairs, underlying their differences (max. 2 pages).

2. Fit a linear model on the entire dataset, using the variable `mpg` as outcome. Give an estimate of the standard error or each estimated coefficient using Bootstrap. Use both bootstrapping of pairs and of residuals.

3. Compute the classical t-distribution confidence intervals, and the percentile Bootstrap intervals for all model coefficients. Use both bootstrapping of pairs and of residuals. If running the code gives you an error, try to explain why does it happen (note that it will not necessarily happen!).

4. We are now interested in fitting a model using only the variables `wt`, `qsec`, and `am`. Re-estimate the model and compute all previous confidence intervals. Comment on the differences between the obtained intervals (if any).

5. **Bonus.** Perform a test on the effect of the variable `am` based on Bootstrapping the residuals of the null model. Compute the p-value of the test, and compare it with the classical t-test p-value. Comment on the result.