# Multiple Linear Regression in R

# (predicting income – train and test)

This exercise uses file Income\_data\_categ.csv.

Before a predictive model is used, it has to be tested, and its accuracy evaluated.

## Tasks to be performed:

1. Split the available data into training and testing (hold out) set in ratio 80%:20%.
2. Build the predictive model of the dependence of income on age, education, smoking and residence. Use only the training set.
3. Find out if one of the independent variables is insignificant and rebuild the model without it.
4. Compare the quality of the model (explanatory power) and see if removing that variable made the model any worse.
5. Compute the predictions for the testing set.
6. Calculate the model error using Mean Absolute Percentage Error (MAPE):

Where:

n – number of data points in the testing set

**–** actual value of the dependent variable (income) for testing data point ***i***

**–** predicted value of the dependent variable (income) for testing data point ***i***