

## **Master BeNeFri in Computer Science**

Course: Statistical Learning Methods  
Spring 2017

### **Exercise #1: Manipulation of data with R**

Download from the ILIAS website the dataset Education dataset (filename: Education.txt) and read the description of the data given in the pdf file "Education.pdf".

1. For each of the variable, compute the mean, the median, the standard deviation, the minimum and maximum value.
2. Select the variables according to the underlying country. For each country, compute the mean, the median, the standard deviation, the minimum and maximum value. Do you see some differences between the countries?
3. (Ignore the difference between the countries). What do you think / infer from all these variables when the main focus is to predict the value of the variable Wage?
4. As the variable Gender is a categorical (binary data), select the wage and education values corresponding to each of the two possible gender values. Compute the mean, the median, the standard deviation, the minimum and maximum value for each gender separately. What can you infer from these values?