

Master BeNeFri in Computer Science

Course: Statistical Learning Methods with R
Spring 2022

Exercise #1: Data Manipulation with R

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

1. For each variable, compute the mean, the median, the standard deviation, the minimum and maximum value.
2. 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?
3. Can you detect some inconsistencies in the data? If yes, what would you do to handle them?
4. What can you infer from all the other variables when the main focus is to predict the values for `Wage`?