Master BeNeFri in Computer Science

Course: Statistical Learning Methods with R
Spring 2022

Exercise #2: Plot() and Statistical Reasoning

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

1. Using the plot() function, show *graphically* the possible relationship between the independent variable `Education` and the dependent variable `Wage`. Can you change the main title of the plot as well as the labels appearing in the two axes?

2. As the variable `Gender` is a factor (categorical / binary data), select the `Wage` and `Education` values corresponding to each of the two possible `Gender` values. Show graphically the possible relationship between the independent variable `Education` and the dependent variable `Wage` but only when considering the observations corresponding to *men* then with *women*. Can you also plot on the same graphics the information for *men* and *women* (with two different colors)?

Download from the ILIAS website the dataset coming from the Simpson's hospital (filename: Simpson.txt) and read the description of the data given in the file "Simpson.pdf".

3. A new patient is coming and the Simpson's hospital diagnoses a tumor in this person. Considering only the variables `Result` and `Treatment` what do you suggest as the most promising treatment for this person?

4. The Simpson's hospital discovers that the size of the tumor is small for this patient. Do you change your suggestion?

5. And if the Simpson's hospital discover that the size of the tumor is large, do you change your conclusion?

6. Are you consistent in the two cases above? If no, why?