Master BeNeFri in Computer Science

Course: Statistical Learning Methods with R
Spring 2022

Exercise #4: Linear regression with R

Download from the ILIAS website the dataset “EducationBis” (filename: EducationBis.txt), containing a new version of the “Education” data.

1. Build two linear models, one for men and one for women, and use the `Education` variable to explain the `Wage`. Describe the results of the output provided by the function `lm()` for both models.

2. Are the two slopes significantly different from 0?

3. Can you build a simple `lm()` model using all the predictors? Describe this unified model.

Download from the ILIAS website the dataset “Computers” (filename: Computers.txt) and read the description of the data given in the file “Computers.pdf”. The performance of the system (response) is indicated either by the variable `ERP` or `PRP`. We will use the target variable `PRP`.

4. Check the different variables (predictor) you have to predict `PRP`. In your opinion, which are the variables that cannot be used to explain the system performance?

5. You’re allowed to use only a single variable (predictor) to predict the value of `PRP`. Which one would you use? Does your model explain something? What is the confidence interval around the slope?

6. Visualize graphically the (linear) relationship that you found.

Download from the ILIAS website the dataset “Cars” (filename: Cars.txt) and read the description of the data given in the file “Cars.pdf”. The performance of the system (response) is indicated by the variable `mpg`.

7. Check the different variables (predictor) you have to predict `mpg`. In your opinion, which are the variables that cannot be used to explain the system performance?

8. You’re allowed to use only a single variable (predictor) to predict the value of `mpg`. Which one would you use? Does your model explain something? What is the confidence interval around the slope?

9. Visualize graphically the (linear) relationship that you found.