

Information Mining - winter semester 2020**Exercise sheet 6**

Exercise 1: R and RStudio

R ¹ is a tool for statistic analysis and graph-generation. A GUI for R is *RStudio*.² Install R and *RStudio* and get to know the Application. A installation guide and a help are available on the given websites.

Exercise 2: Statistics with R

Calculate the following values using the example data³:

- (a) Median of the age
- (b) Arithmetic mean of the earnings
- (c) Number of married people

Draw a histogram with the data from the age column. The histogram should contain 6 intervals.

Exercise 3: Significance test and correlation with R

Use the example data from exercise 2.

- (a) Is there a significant difference between the earning of people who responded and those who did not respond? Use the t-Test to calculate.
- (b) Does a correlation between the age and the earning exist? Make a correlation analysis according to Spearman.

(Please turn the page)

¹<http://www.r-project.org>

²<http://www.rstudio.com/products/RStudio/#Desktop>

³dataA16.csv

Exercise 4: Linear regression with *R*

The example data⁴ contain the number of new students in the last years.

- (a) Create a scatter plot of the data
- (b) Calculate the regression line and draw it into the plot
- (c) The linear regression can be used as a prediction method. How many students will be expected for the year 2016?
- (d) Does a correlation between the year and the number of new students exist? Make a correlation analysis according to Spearman.

⁴dataR.csv