First Exam

UTB

Data Science

Exercise 1

Generate two vectors with the ages (in years) and the heights (in cm) of 20 persons by means of the following syntax:

```
age <- rpois(20, 40)
hei <- round(rnorm(20, 175, 10), 2)
```

- a) What is the height of the tallest person?
- b) Which is the person with the lowest height and how old is that person?
- c) How many people have an age between 30 and 40 years, both including?
- d) Use function sample to create a vector called uni with the university careers of these people: Statistics, Physics, Biology, and Medicine.
- e) What is the median age of those who studied physics or medicine?
- f) Overwrite the values of 'Statistics' and 'Physics' by 'Stats' and 'Phys', respectively.
- g) Create a data frame that contains the three variables —age, hei, uni— and save the workspace.

Exercise 2

The object state.x77 of package datasets contains data (recorded in 1977) of the 50 states of the United States whereas the vector state.region is a factor that contains the corresponding regions. To display both objects you can execute the following instructions in R:

```
state.x77
state.region
```

- a) Is object state.x77 a matrix or a data frame?
- b) How many variables does state.x77 have?
- c) Convert state.x77 into a data frame with name states77 and add state.region as a new variable with name Regio.
- d) Delete the variables HS.Grad and Frost.

Execute the following command in order to change the variables' names (function tolower converts them in lower case and function substr reduces them to the first four letters):

```
names(states77) <- substr(tolower(names(states77)), 1, 4)
```

- e) Overwrite the life expectancies of the first and last state, Alabama and Wyoming, by the values 71.3 and 72.1, respectively.
- f) Compute the correlations (of Pearson) among the variables inco, illi, and murd.
- g) Which of the four regions has the highest average salary? How much is it?

Exercise 3

- a) Import the data of file Test1Exercise3.txt to R using function read.table.
- b) Which are Pedro's age, height, and weight?
- c) Calculate the mean of variable 'Height' and the median of variable 'Weight'.