

DK.0016eng. Homework
(autumn semester 2021)

Max result: 20 points

The following exercises base on the University of Tartu students' dataset:

http://ph.emu.ee/~ktanel/DK_0016/students.RData.

You can import this dataset into R with command

```
load(url("http://ph.emu.ee/~ktanel/DK_0016/students.RData"))
```

The values of some of the variables used in the following tasks are as follows:

gender: 1 – woman, 2 – man; sport: 1 – not doing sports, 2 – doing workout 1..2 times a week, 3 – doing workout 3..4 times a week, 4 – doing workout 5 or more times a week; health: 1 – very good, 2 – good, 3 – moderate, 4 – bad; ...

1. Calculate the students' body mass index (as a new variable) and study its distribution separately for men and women. Does the body mass index follow the normal distribution?
2. What is the average (\pm standard deviation) body mass index of men and women, Is the difference statistically significant? Make the decision both based on the 95% confidence interval of means' difference and based on the p-value.
3. Create a new variable 'sport01' with value zero, if student does not practice sport (sport=1), and value one, if student practices sport (sport>1). You can use the following command:

```
sport01 <- factor(sport, levels=1:5, labels=c(0,1,1,1,1)).
```


How big is the percentage of students practicing sport (with 95% confidence interval)? Is the percentage of students practicing sport statistically significant from 75%?
4. What about the percentage of sporting students among men and women? Are these proportions statistically significantly different?
5. Are the body mass index and sporting associated? Test the statistical significance of body mass index difference between sporting and non-sporting students. Visualize the result
6. Are the sporting and health associated? Test, visualize and comment the result.
7. Are the beer consumption and sporting of male students associated?
8. Create a new variable 'smoke012' with value zero, if the student does not smoke (smoking=1), one, if the student no longer smokes but has smoked (smoking=2), and two, if the student smokes (smoking>2). Does the smoking depend on gender? Visualize the result.
9. Are student body mass index and systolic and diastolic blood pressure related? Whether these relationships are different between men and women (no statistical significance testing of difference is required, but you can visualize the results)?
10. Predict students' systolic blood pressure based on gender and body mass index. Illustrate the result of the modeling. What is the expected systolic blood pressure of male and female students with a body mass index of 20 points? But with body mass index of 25 points? Is it necessary to consider also the gender by body mass index interaction? But sporting and/or smoking?