mlcourse.ai. Assignment #2 (demo)

Total points 18/18

Analyzing cardiovascular disease data

✓ 1. How many men and women are present in this dataset? Values of the gender feature were not expailned (whether "1" stands for women or men) – figure this out by looking analyzing height, reasonably assuming that on average men are taller. 1/1

- (●) 45530 women and 24470 men ✓
- ( ) 45530 men and 24470 women
- ( ) 45470 women and 24530 men
- ( ) 45470 men and 24530 women

✓ 2. Who more often report consuming alcohol - men or women? 1/1

- ( ) women
- (●) men ✓

✓ 3. What's the rounded difference between the percentages of smokers among men and women? 1/1

- ○ 4
- ○ 16
- ◉ 20 ✓
- ○ 24

✓ 4. What's the rounded difference between median values of age for non-smokers and smokers? You'll need to figure out the units of feature age in this dataset. 1/1

- ○ 5
- ○ 10
- ○ 15
- ◉ 20 ✓

✓ 5. Calculate fractions of ill people (with CVD) in two segments described in the task. What's the quotient of these two fractions? 2/2

- ○ 1
- ○ 2
- ◉ 3 ✓
- ○ 4

✓ 1.6. Choose the correct statements 2/2

- [ ] Median BMI in the sample is within boundaries of normal values.
- [x] Women's BMI is on average higher then men's. ✓
- [ ] Healthy people have, on average, higher BMI than ill people.
- [x] In the segment of healthy and non-drinking men BMI is closer to the norm than in the segment of healthy and non-drinking women ✓

✓ 1.7. How many percents of data (rounded) did we throw away? 2/2

- ○ 8
- ○ 9
- ◉ 10 ✓
- ○ 11

✓ 2.1. Which pair of features has the strongest Pearson's correlation with gender feature? * 2/2

- ○ Cardio, Cholesterol
- ◉ Height, Smoke ✓
- ○ Smoke, Alco
- ○ Height, Weight

✓ 2.2. Which pair of features has the strongest Spearman's correlation between each other? 2/2

- ○ Height, Weight
- ○ Age, Weight
- ○ Cholesterol, Gluc
- ○ Cardio, Cholesterol
- ◉ Ap_hi, Ap_lo ✓
- ○ Smoke, Alco

✓ 2.3. Why do these features have strong rank correlation? 2/2

- ○ Inaccuracies in the data (data acquisition errors)
- ○ Relation is wrong, these features should not be related
- ◉ The nature of the data ✓

✓ 2.4. What is the smallest age at which the number of people with CVD outnumber the number of people without CVD? 2/2

- ○ 44
- ◉ 55 ✓
- ○ 64
- ○ 70

Do you have any remarks concerning the assignment? In case of apparent errors/typos please use GitHub Issues and/or Pull Requests (https://github.com/Yorko/mlcourse.ai).