IDS2018 - Assignment 1

Isabela Blucher

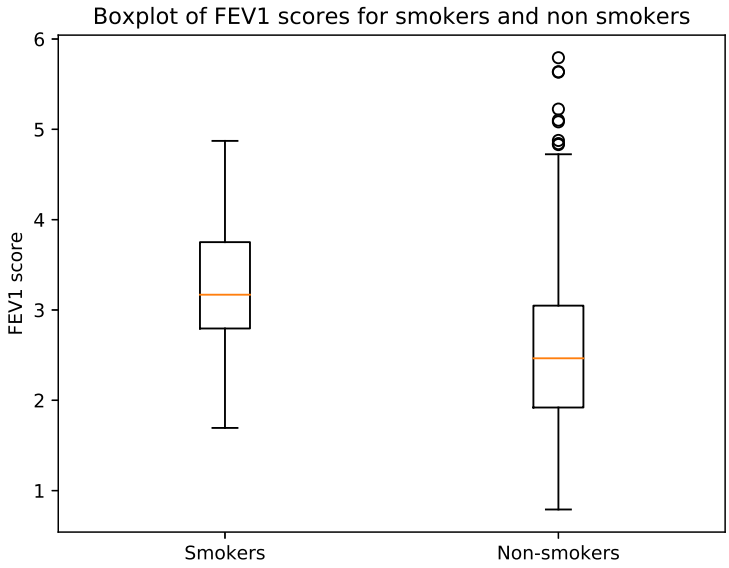
February 15, 2018

Exercise 1

Having read and separated the data, I computed the average FEV1 scores for both smokers and non-smokers. The computed average for smokers was 3.27686153846 and for non-smokers it was 2.5661426146. The results are surprising because a lower FEV1 score indicates a decrease in lung function and yet the smokers have a higher average score than the non-smokers.

Exercise 2

The inner-quartile range for the smokers has a larger FEV1 score, which is surprising given that the expected is that the non-smokers would have better lung function. One indication of good lung function for the non-smokers would be the large upper whisker and the outliers of its boxplot.



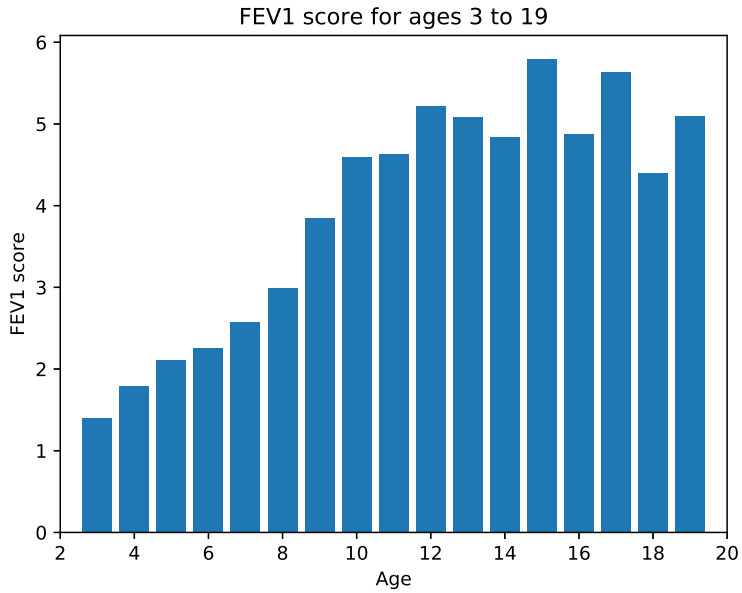
Exercise 3

After the implementation of the two-sided t-test for hypothesis testing, the values computed were for the t-statistic = 7.14961, for the degrees of freedom $\nu = 84$ and for the returned p-value = 2.96184e-10. Since the returned p-value is smaller than the given significance level of 5%, there is enough evidence to reject the null hypothesis. This implies that there is a statistically significant difference between the FEV1 scores of smokers and non-smokers.

Exercise 4

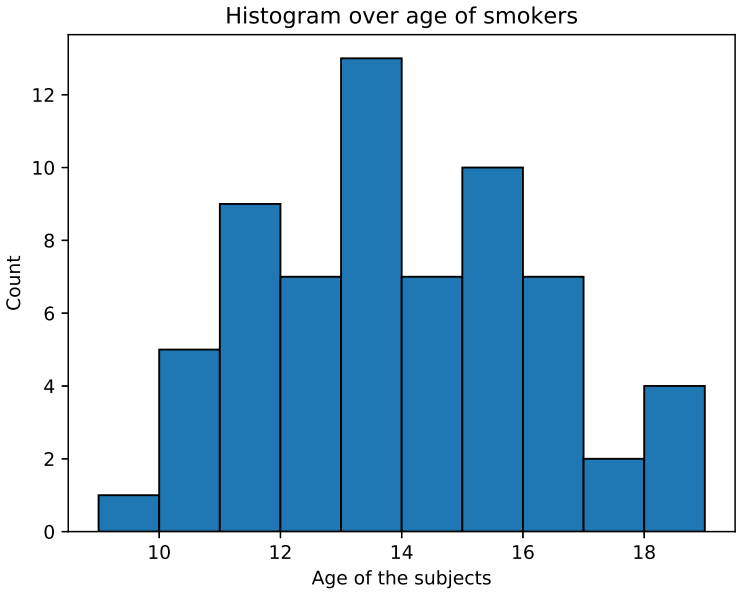
The bar plot below illustrates that the older a child gets, the more the FEV1 score increases. This can be verified by the correlation coefficient which was computed as r = 0.75645898999 for the

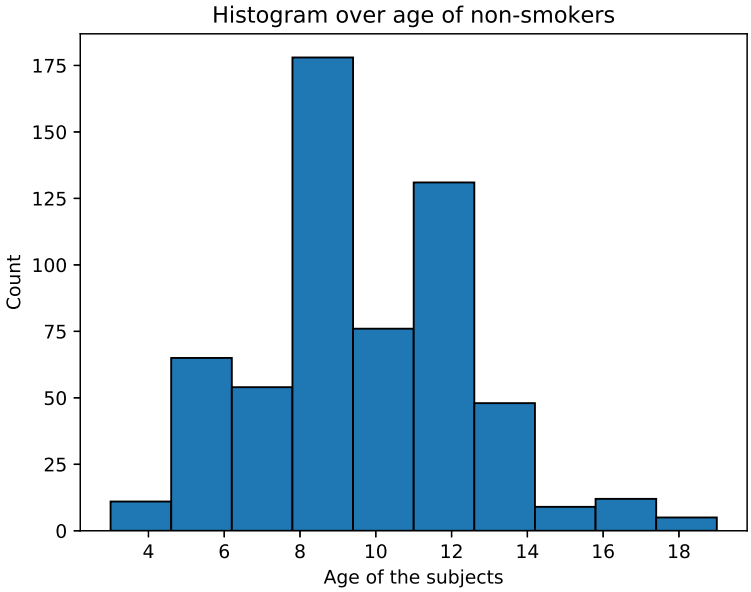
given dataset. This value implies positive correlation between a child's age and its FEV1 score.



Exercise 5

The histograms below illustrate that the smokers are generally older than the non-smokers. Most smokers are 11 and up, while most non-smokers are kids between the ages of 3 and 10. This would explain the obtained results on lung function that smokers have better average FEV1 score than non-smokers, because the older the kid, the better the FEV1 score, and consequently, the lung function.




Histogram over age of non-smokers
Count
0
25
50
75
100
125
150
175
4
6
8
10
12
14
16
18
Age of the subjects