

# Stefano Pellegrini - Assignment 1

February 16, 2020

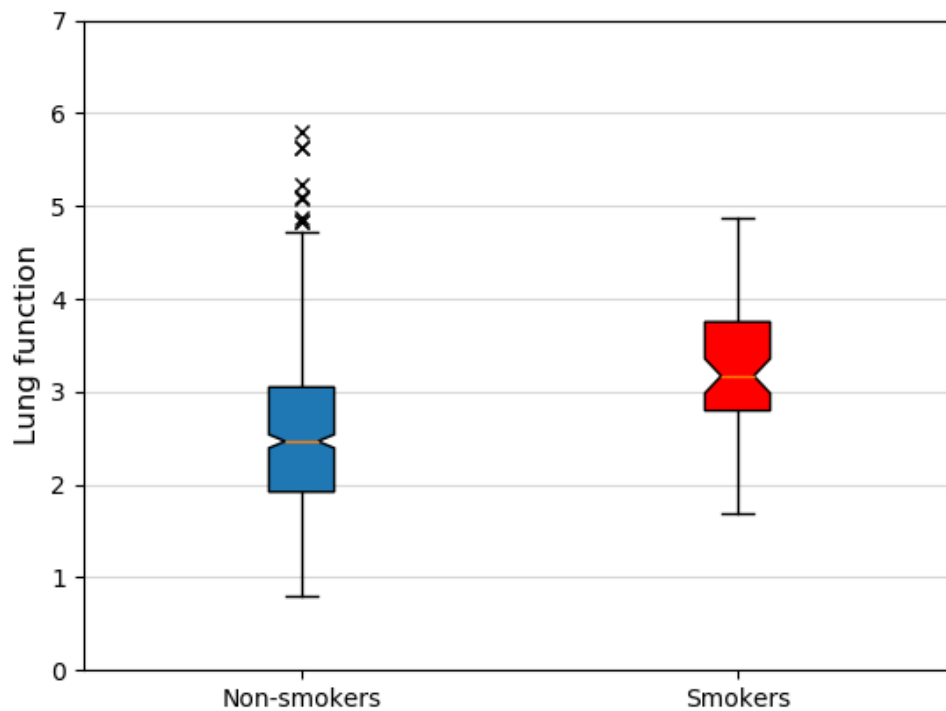
## Exercise 1

- Average lung function among non-smokers = 2.5661426146010187
- Average lung function among smokers = 3.2768615384615383

Surprisingly, smokers have a higher average lung function than non-smokers.

## Exercise 2

Boxplot over lung functions for non-smokers and smokers



I can observe that in non-smokers several observations are outside the whisker range and have quite large lung function. Also, the median and the interquartile range of the lung function for this group is lower than the lung function median and interquartile range for smokers. Furthermore, the

non-smokers have a smaller minimum compared to the smokers, while there isn't a big difference between their maximum.

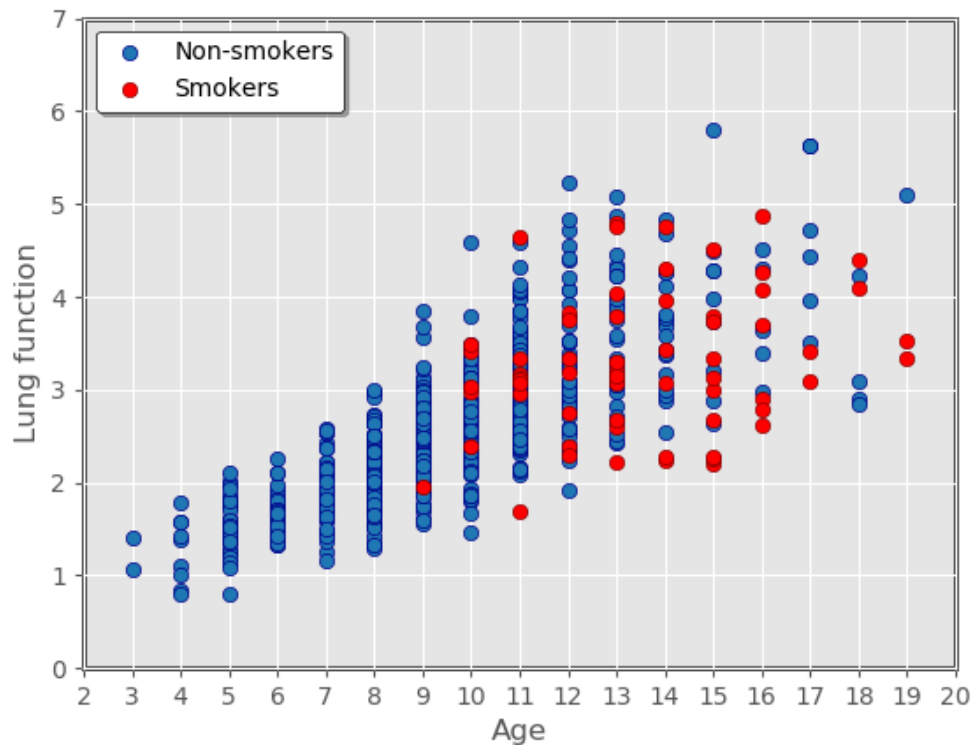
### Exercise 3

- T-statistics = -7.1990318609997095
- Degrees of freedom = 83
- P-value = 2.4945644815274697e-10
- Reject the null hypothesis = True

The p-value is very small and I reject the null hypothesis at a significance level alpha of 0.05. So I accept the alternative hypothesis that the mean of the lung function of the smokers' population is different from the mean of the lung function of the non-smokers' population and, that the observed difference between the two samples mean is not given by chance.

### Exercise 4

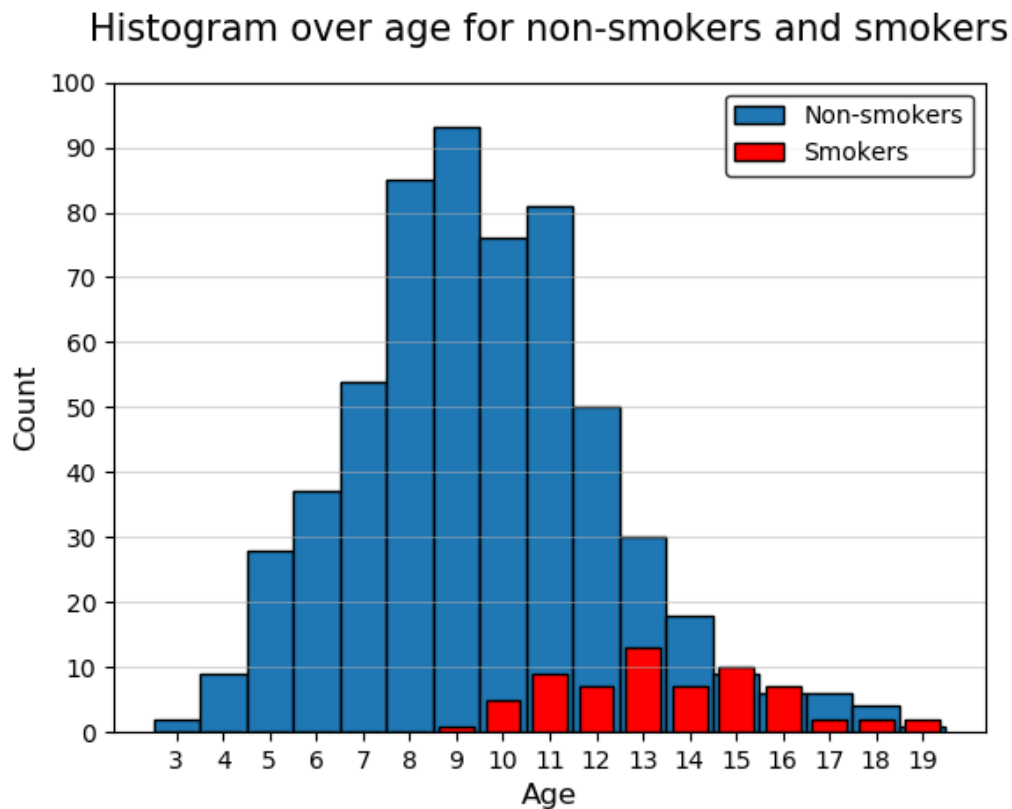
2D plot of lung function over age for smokers and non-smokers



- Pearson (all data) = 0.7564589899895999
- Spearman (all data) = 0.7984229001546537
- Pearson (non-smokers) = 0.7815977186520366
- Pearson (smokers ) = 0.24907515552578594
- Spearman (non-smokers) = 0.8123113987207394
- Spearman (smokers) = 0.22808903187369514

The correlation coefficients show that there is a general positive correlation between age and lung function. Also, dividing the dataset between smokers and non-smokers we can observe that, although the correlation is positive in both groups, in the non-smokers the relationship is strong, while in the smokers it is weak. In the 2D plot is possible to observe that in the non-smokers there is a positive linear relationship between age and lung function, but, we can also observe that there are several non-smokers that are much younger than the youngest smoker. So the main reason why we observed such a larger correlation, in non-smokers compared to smokers, is not because of smoking but because there is a big difference in the age distribution between the two groups.

## Exercise 5



As previously observed, the subjects in the smokers group are older than the non-smokers. Because lung and thorax development continue at least during puberty (Nève et al.), looking at the age distribution of the two groups, it is easy to understand that the smokers have a larger lung function because they are all older than 8 years old. Also, it is possible to observe that our data set contains much more observations of non-smokers than smokers.

## Reference

V. Nève, F. Girard, A. Flahault, M. Boulé. Lung and thorax development during adolescence: relationship with pubertal status. *European Respiratory Journal* 2002 20: 1292-1298