# Homework Week 8

### Problem 1:

Download the Boston Housing Data Set from Kaggle. (https://www.kaggle.com/arslanali4343/real-estate-dataset). A convenience copy is under Week8Exercises. The meaning of the columns are as follows:

- CRIM - per capita crime rate by town
- ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
- INDUS - proportion of non-retail business acres per town.
- CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
- NOX - nitric oxides concentration (parts per 10 million)
- RM - average number of rooms per dwelling
- AGE - proportion of owner-occupied units built prior to 1940
- DIS - weighted distances to five Boston employment centres
- RAD - index of accessibility to radial highways
- TAX - full-value property-tax rate per $10,000
- PTRATIO - pupil-teacher ratio by town
- B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
- LSTAT - % lower status of the population
- MEDV - Median value of owner-occupied homes in $1000's.

Import this as a pandas data frame. Use describe on the data frame.

### Problem 2:

Determine with all tests presented whether the PTRATIO and MEDV for the data set are normally distributed.

### Problem 3:

Test whether CHAS "bounding the Charles River" and MEDV are related assuming that MEDV is normally distributed. With other words: Are the means of the MEDV different depending on whether the property bounds the Charles River or not.

### Problem 4:

The following gives a (probably made-up statistics on the preferences of cyclists. Use the $\chi^2$ test to decide whether there are differences between the genders.

| | Lake Path | Hill Path | Wooded Path |
|---|---|---|---|
| **female** | 45 | 38 | 27 |
| **male** | 26 | 52 | 12 |

Use chi2_contingency from scipy and interpret the result. Repeat for the "log likelihood test" (see the documentation of chi2_contingency).