Statistics with Python
Practicals 3

Data manipulating

Exercise 1

☞5 points

- Read in data from *experiment.xls* file (**Hint**: see for instance `http://pandas.pydata.org/pandas-docs/stable/generated/pandas.read_excel.html`), which contains data from an experiment with plants in three different growing conditions.
- Create a *DataFrame* and extract the relevant columns in order to group the weight data by the 'group' column. Use the method *describe()* and print its output table.
- Plot a *boxplot* for the three groups (control, treatment A and treatment B). What are the values you can identify on the *boxplot* that are also given in the *describe()* table?
- Create another *DataFrame* from the read-in data, that contains only three columns: 'control', 'treatmentA', and 'treatmentB'. For these three columns, extract only the data from lines 0 to 9.
- Extract the values from the column 'control' into an *array*. Plot these values into a *scatter* plot. Plot the mean of these values with a dashed horizontal line on the scatter plot.

Hypothesis tests

Exercise 2

☞1 point
Paired t-Test and Wilcoxon signed rank sum test
The daily energy intake from 11 healthy women is [5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, 8770] kJ. Is this value significantly different from the recommended value of 7725?

Exercise 3

☞3 points
Two Groups
In a clinic, 15 lazy patients weigh [76, 101, 66, 72, 88, 82, 79, 73, 76, 85, 75, 64, 76, 81, 86.] kg, and 15 sporty patients weigh [64, 65, 56, 62, 59, 76, 66, 82, 91, 57, 92, 80, 82, 67, 54] kg.

- **Normality test**: Are the two data sets normally distributed?
- **t-Test of independent samples**: Are the lazy patients significantly heavier?
- **Mann-Whitney test**: Are the lazy patients still heavier, if you check with the Mann-Whitney test? (Hint: try two-tailed and one-tailed tests. The default is a one-sided p-value.)

Sensitivity and specificity

Exercise 4

☞1 point
We consider two pregnancy tests. If test 1 gives TP=980 and FN=20, and test 2 gives TP=2800 and FN=200, which of the tests is better and why?

☞The answers to the questions above must be handled in today, at the end of the practicals session!