Theoretical Exercises

Excercise 5.1: (Theoretical) Central tendencies and statistical deviation

In a survey, send to several companies, the question

"How large have been your expenses (in thousand Euros) in the last year?"

was asked. 11 companies answered it with the following result:

company	1	2	3	4	5	6	7	8	9	10	11
expenses	110	75	70	65	55	70	140	90	90	55	90

a) Calculate the arithmetic mean, the range and the empirical standard deviation $\tilde{s}$.

b) Calculate and draw the empirical distribution function. Then, determine the upper quartile, lower quartile and median. Draw a box-plot.

c) Derive a linear transformation $y_i = ax_i$, such that the empirical variance of y_i equals to 1.

Excercise 5.2: (Theoretical) Small proofs

Let $x_1, \ldots, x_n$ be observations of a metric variable X. Show that the following equations, given in the lecture, hold:

- $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$
- $\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2$

Practical Exercises

Excercise 5.3: (Practical) Frequencies of data

Consider the Ames housing dataset given in Exercise 2. Load the dataset with the help of pandas and perform the following tasks:

a) Compute the median, mean, both quartile and both decile of the variable ”SalePrice”.

b) Compute the range, quartile difference and empirical standard deviation. Compare these values and decide which value seems to be more precise, i.e. describes the smallest range.

c) Create box-plots of the variables ”Year Built” and ”Year Remod/Add”. Compare both plots.