Statistics and Sensor Data Fusion
– Winter Term 2023/2024 –
Worksheet 2
Prof. Dr.-Ing. Gernot Fabeck

Exercise 1. The table below presents data belonging to eight different high-tech companies considering the size of each company (number of employees) and the average duration of employment (in years). Calculate the corresponding **coefficients of correlation of Spearman** and **Bravais-Pearson** and interpret the values of the two coefficients:

number of employees	221	251	346	376	401	421	471	481
average duration of employment	9.7	7.9	8.6	7.2	7.3	7.1	7.0	6.8

Exercise 2. An enterprise focuses on the production of only one kind of product (integrated assembly lines). Over a period of five seasons, production volume and total production costs (in million EUR) have been recorded as follows:

number of produced units	9	12	14	12	13
total costs of production	12	13	14	13	13

Compute the **fixed costs** of the enterprise (costs that do not depend on the amount of goods or services produced by the enterprise) and try to indicate the **variable costs** (costs that change in proportion to the amount of produced goods or services). Assess the quality of your regression analysis by means of the **coefficient of determination**.

Exercise 3. The following two-dimensional raw data has been recorded:

$$\left((x_i, y_i)\right)_{i=1}^{4} = \left((1,4), (5,120), (7,253), (9,418)\right)$$

Inspection of the data and previous knowledge propose a **regression function** of the type $y = f(x) = b \cdot x^a$, $a, b \in \mathbb{R}$. Determine the regression function based on this suggestion.

Exercise 4. Right or wrong?

(a) The raw data can always be reconstructed from a table of frequencies.

(b) The coefficient of determination improves with the number of observed values.

(c) In linear regression analysis, the point $(\bar{x}, \bar{y})$ always hits the trend line.

(d) For every regression line, it holds for the residuals $\hat{u}_i$ that

$$\sum_{i=1}^{n} \hat{u}_i = 0$$