

## A mathematical method to find the „best“ curve: **the least squares method**

We want to find the values of the parameters  $\alpha, \beta, \dots$  in  $y = f(x, \alpha, \beta, \dots)$  by minimizing the sum

$$S(\alpha, \beta, \dots) = \sum_{i=1}^N (f(x_i, \alpha, \beta, \dots) - y_i)^2.$$

We have  $N$  pairs of measured data  $x_i, y_i$  ( $i = 1, 2, \dots, N$ ).

The conditions of the minimum:

$$\frac{\partial S}{\partial \alpha} = 0, \quad \frac{\partial S}{\partial \beta} = 0, \dots$$

By solving this algebraic equation system we can get the best values for  $\alpha, \beta, \dots$ .

This method can be applied when

- the error of  $x$  is negligible compared to the error of  $y$ , and
- the variance of  $y$  is independent of the value of  $x$ .

In case of a **linear function**  $y = ax + b$

the sum of squares of the differences is  $S(a, b) = \sum_{i=1}^N (ax_i + b - y_i)^2$ ,  
so we get the following algebraic equations:

$$\frac{\partial S}{\partial a} = 2 \sum_{i=1}^N (ax_i + b - y_i) \cdot x_i = 0 \Rightarrow a \sum_{i=1}^N (x_i^2) + b \sum_{i=1}^N (x_i) - \sum_{i=1}^N (x_i \cdot y_i) = 0$$

$$\frac{\partial S}{\partial b} = 2 \sum_{i=1}^N (ax_i + b - y_i) \cdot 1 = 0 \Rightarrow a \sum_{i=1}^N (x_i) + N \cdot b - \sum_{i=1}^N (y_i) = 0$$

The solution is

$$b = \frac{\sum_{i=1}^N (y_i)}{N} - a \frac{\sum_{i=1}^N (x_i)}{N} \quad \text{and} \quad a = \frac{\sum_{i=1}^N (x_i \cdot y_i) - \frac{\sum_{i=1}^N (x_i) \cdot \sum_{i=1}^N (y_i)}{N}}{\sum_{i=1}^N (x_i^2) - \frac{(\sum_{i=1}^N (x_i))^2}{N}}.$$

By introducing the averages

$$\bar{x} = \frac{\sum_{i=1}^N (x_i)}{N}, \quad \bar{y} = \frac{\sum_{i=1}^N (y_i)}{N}, \quad \overline{x \cdot y} = \frac{\sum_{i=1}^N (x_i \cdot y_i)}{N}, \quad \overline{x^2} = \frac{\sum_{i=1}^N (x_i^2)}{N}$$

the solution can be written as

$$a = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \quad \text{and} \quad b = \bar{y} - a \cdot \bar{x}$$

In case of a **linear function**  $y = ax$  (the intercept of the line is zero)

$$S(a) = \sum_{i=1}^N (ax_i - y_i)^2 \rightarrow \min. \quad \text{leads to} \quad a = \frac{\overline{x \cdot y}}{\overline{x^2}}.$$

The **standard deviation**

of the slope and of the intercept can be calculated applying the following formulae:

In case of  $y = ax + b$

$$s_a = \sqrt{\frac{\sum_{i=1}^N (a \cdot x_i + b - y_i)^2}{N \cdot (N-2) \cdot (\bar{x}^2 - \bar{x}^2)}} \quad \text{and} \quad s_b = \sqrt{\bar{x}^2} \cdot s_a$$

In case of  $y = ax$

$$s_a = \sqrt{\frac{\sum_{i=1}^N (a \cdot x_i - y_i)^2}{N \cdot (N-1) \cdot \bar{x}^2}}$$