The method of least squares

This method is applicable when we measure two quantities bound by some functional dependence. Let us assume that we obtained pairs of measured data (x_1, y_1) … (x_n, y_n), which represent points in the *xy* plane. Our goal is to find the dependence

$$y = f(x),$$

where x is the independent variable and y is the dependent variable and n is number of measurements. The function *f* is called *regression function* which is unequivocally determined by its parameters. The most frequently used method for the parameter determination is the least squares method deduced by Gauss.

The principle of the least squares method is finding such parameters of the *f* function, so that the sum of squares of differences between measured and calculated data is minimal. In another words, the expression

$$q = \sum_{i=1}^{n} [y_i - f(x_i)]^2 \qquad [1]$$

must be minimal. The most simple type of dependence is the linear one represented by a line equation

$$f(x) = y = ax + b \qquad [2]$$

Substituting formula [2] into [1] we obtain

$$q = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

To be able to find *a* and *b* parameters for which the *q* is minimal, we have to use partial derivations

$$\frac{\partial q}{\partial a} = -2x_i \sum_{i=1}^{n} (y_i - ax_i - b) = 0 \qquad [3]$$

$$\frac{\partial q}{\partial b} = -2\sum_{i=1}^{n} (y_i - ax_i - b) = 0 \qquad [4]$$

By a small arrangement we obtain a couple of *normal equations*

$$b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^{\ 2} = \sum_{i=1}^{n} x_i y_i \qquad [5]$$

$$nb + a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad [6]$$

Evaluating from the set of equations [5] and [6] we can obtain required parameters

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad [7]$$

$$b = \frac{\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n} \qquad [8]$$

Taking into account the definition of average value

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

we can simplify equations [7] and [8] into

$$a = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad [9]$$

$$b = \bar{y} - a\bar{x} \qquad [10]$$

For the uncertainty determination we can use the following formulae (without deduction):

$$s_a = \frac{s}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \qquad [11]$$

$$s_b = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \qquad [12]$$

Where $$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} [y_i - \bar{y} - a(x_i - \bar{x})]^2}$$

The s_a is uncertainty of the *a* parameter (slope of the line) while s_b is uncertainty of the *b* parameter (a constant representing intersection with the *y* axis).