Linear regression

Linear regression allows to determine the parameters of the linear function $y = ax + b$ best fitted to the measurement points if there is a linear relationship between them (or if it can be brought to such a dependence by means of a corresponding transformation). After comparing the equation of straight line with a theoretical dependence between the measured quantities, it is possible (on the basis of the calculated values of the directional coefficient *a*, the free expression *b* and their measurement uncertainties Δa and Δb) to determine the physical quantities sought in the experiment together with their measurement uncertainties.

Method 1

1. Based on measured experimental values x_i and y_i we complete the following table:

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1					
2					
3					
…					
n					
	$\sum x_i = \ldots$	$\sum y_i = \ldots$	$\sum x_i^2 = \ldots$	$\sum y_i^2 = \ldots$	$\sum x_i y_i = \ldots$

2. We calculate the values of the directional coefficient *a* and the free expression b:

$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}, \qquad b = \frac{\sum y_i - a\sum x_i}{n}.$$

3. We calculate the values of measurement uncertainty of the determined coefficients a and b

$$\Delta a = \sqrt{S^2 \cdot \frac{n}{n\sum x_i^2 - \left(\sum x_i\right)^2}}, \qquad \Delta b = \sqrt{S^2 \cdot \frac{\sum x_i^2}{n\sum x_i^2 - \left(\sum x_i\right)^2}}$$

$$\text{gdzie: } S = \sqrt{\frac{\sum y_i^2 - a\sum x_i y_i - b\sum y_i}{n-2}}.$$

4. We calculate the linear correlation coefficient R:

$$R = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left(n\sum x_i^2 - \left(\sum x_i\right)^2\right)\left(n\sum y_i^2 - \left(\sum y_i\right)^2\right)}}.$$

Method 2

In order to determine the linear regression coefficients, you can also use Excel:

Example 1:

1. We introduce experimental data into two columns (x_i, y_i).

2. Next to the entered data, we mark six boxes, lying next to each other in two columns and three lines.

3. With the data fields still selected, enter the command line:

=REGLINP(the range of the variable y ; the range of the variable x ;1;1)

4. We approve the command: while holding down the **CTRL** and **SHIFT** keys, press the **ENTER** key.

5. The selected cells will show values of the directional coefficient *a*, the free expression *b*, their measurement uncertainties Δa and Δb, and the square of the correlation coefficient *R* - calculated using the least squares method.

a	b
Δa	Δb
R^2	

Example 2

1. We introduce experimental data in two columns (x_i, y_i).

2. We mark the entered data and insert the dot plot only with markers.

3. Click on any point in the data series and select "Add trend line ..." from the expanded menu.

4. We mark "linear" as the trend line type and check the "Display equation in the chart" and "Display R-square in the chart" boxes.

5. We click on the "Close" button. The regression equation and R^2 will appear on the graph.

Attention 1: This method does not allow to determine measurement uncertainties

Δa and Δb

Attention 2: In addition to linear regression, it is also possible to determine the equation of the trend line type: exponential, logarithmic, polynomial, power and moving average.

Method 3

Many calculators allow you to automatically calculate regression coefficients. The description of the procedure, appropriate for a given calculator model, can be found in the calculator's operating instruction.

Method 4

There are many shareware and freeware programs available on the Internet that allow you to set regression coefficients. A simple program in xls format can be found in the "Physics-labs" tab on the website of the KFiCH AM in Szczecin.

Example:

The current I flowing through a resistor of the resistance R as a function of the applied voltage U was measured:

U $[\mathrm{V}]$	I $[\mathrm{mA}]$
1	20.1
3	55.2
5	98.7
7	151.0

After entering the measurement data into the Excel worksheet, selecting six fields and entering the command:

=REGLINP(B1:B4;A1:A4;1;1)

you got:

	A	B	C	D	E
1	1	20,1E-3			
2	3	55,2E-3			
3	5	98,7E-3			
4	7	151,0E-3			
5					
6			0,02181	-0,00599	
7			0,00136	0,006232	
8			0,992285	0,006081	
9					
10					

The sought values from the table are:

$$a = (0{,}0218 \pm 0{,}0014)\,\Omega^{-1}$$

$$b = (-0{,}0060 \pm 0{,}0063)\,\mathrm{A}$$

$$R^2 = 0{,}9922 \;\Rightarrow\; R = 0{,}9961$$

After comparing the equation of the straight line to the theoretical relationship between the measured quantities

$$
\begin{aligned}
y &= a \cdot x + b \\
I &= \frac{1}{R} \cdot U
\end{aligned} \quad ,
$$

it is possible to determine the resistance value R of the resistor and its measurement uncertainty calculated using the exact differential method:

$$
a = \frac{1}{R} \quad \Rightarrow \quad
\begin{aligned}
R &= \frac{1}{a} = \frac{1}{0{,}0218} = 45{,}8715\,\Omega \\
\Delta R &= \frac{\Delta a}{a^2} = \frac{0{,}0014}{0{,}0218^2} = 2{,}9458\,\Omega
\end{aligned}
\quad \Rightarrow \quad R = (45{,}9 \pm 3{,}0)\,\Omega
$$