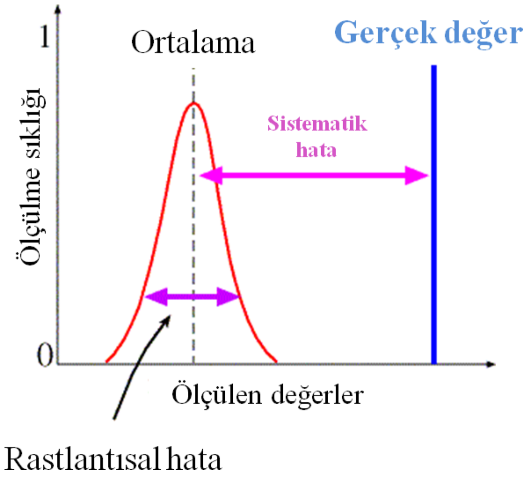
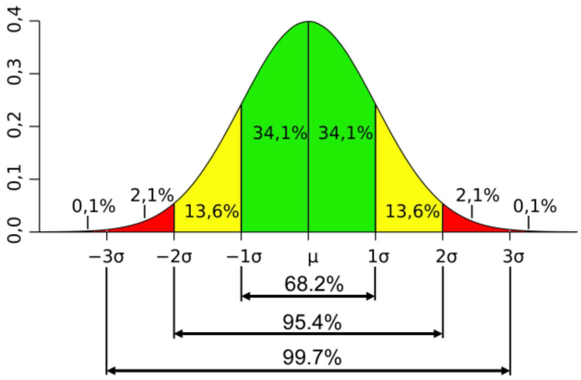
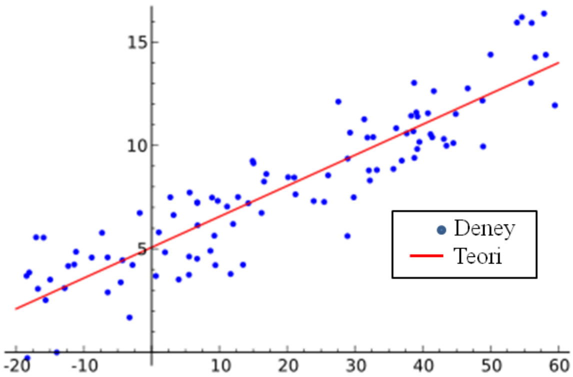
Gebze Technical University
Physics Department

Introduction to Experimental Methods
Basic Concepts, Analysis Methods







Doğan Erbahar

2015, Gebze

This booklet has been formed by expanding and compiling the grades of the 2-hour lecture I have given as an introduction to undergraduate laboratory students at the Physics Department of Gebze Technical University for the past few years.

In the narration, rather than scientific certainty and theoretical setup, presentation ease, contextuality, intellectual integrity, and flow are prioritized. In this context, it was preferred to explain the topics over selected examples rather than abstract and theoretical processing. The details that are likely to be lost and skipped in particular tried to be included in the booklet by highlighting them with footnotes.

Since this is the first version, there are certainly many deficiencies in terms of content, spelling, grammar, and expression, and it is planned to improve in all these aspects over time. In this context, I attach importance to the opinions of readers of all levels and I want them to share these opinions with me.

I would like to thank Ali Kaya, who is one of the laboratories responsible of our department, which emphasizes the necessity of conducting such a study to our university and encourages me in this sense, and my colleagues Yasin Şale, Mehmet Aras, Mustafa Öztürk and Dilek Erbahar, whom I exchanged views during the writing phase.

Doğan Erbahar

03.10.2015

Contents

The importance of experiment in physics

Physics is a branch of science that examines the laws of nature from the most basic level. In terms of conclusions from these investigations, the only decision-making authority is controlled experiments and observations. So much so that a theory that has the slightest disagreement with experiments or observations in a natural event or observation statement remains valid and seeks a new theory. Therefore, controlled experiments and observations are at the forefront of guiding the science of physics.

Measurement

Experiments and observations are carried out through operations called measurement. Measured things are called *physical quantities*. The measure means comparing a physical quantity with its type and a standard accepted unit with the simplest expression and numerically determining how many of those units are hosted. Depending on the sensitivity of each measuring instrument, there is a certain amount of *uncertainty*, and these uncertainties must also be shown and processed when reporting the measurement result.

It is generally possible to divide measurements *directly* and *indirectly*. Direct measurement is carried out by comparing the measured quantity to another quantity of its type. It is an example of measuring the volume of the sphere directly by measuring the volume of a sphere by dipping and increasing the volume of water in a liquid located on a graded cylinder. Another example of measuring directly is that a certain length can be measured by a ruler.

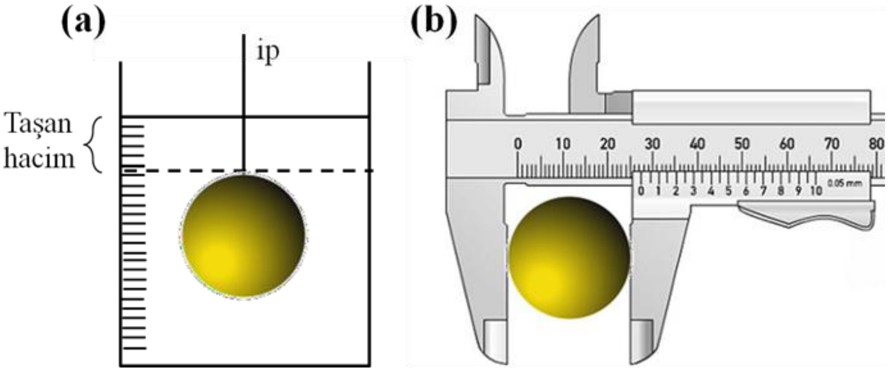


Figure 1 Direct (a) and indirect (b) measurement of the volume of a sphere.

Indirect measurement is based on the measurement of a quantity of "type" other than the quantity that is usually intended to be measured, and from there, the result is reached by accountability. Considering the volume sample of the sphere in the previous paragraph, measuring the diameter with the help of a plot, splitting it in half and finding the radius, and from there, reaching volume using the sphere's volume formula $\frac{4}{3}\pi r^3$ a typical example of indirect measurement.

Now let's discuss some of the concepts we present with italic fonts in the first paragraph:

Physical Quantities

Physical quantities are divided into base and derived quantities. The base quantities adopted in international scientific standards are *mass, distance, time, temperature, amount of substance matter, electrical current, and luminous intensity*. In nature, all other quantities other than these can be derived from them. (For example: speed = length/time, area = length^2, density = mass/length^3, force = mass x length/time^2, etc...)

International System of Units (SI: Système International)

Units and dimension symbols of base physical quantities set by international standards (Table1).

Quantity name	Unit name	Unit symbols	Dimension symbol
Mass	kilogram	kg	M
Time	second	s	T
Length	meter	m	L
Temperature	kelvin	K	T
Electric current	amper	A	I
Amount of substance	mole	mol	N
Luminous intensity	candela	cd	J

Table 1 Base physical quantities.

The derived physical quantity is also combined with the way the typed method is derived within the sample logic at the end of the previous section (For example unit of speed m/s, unit of density kg/m^3, unit of the area m^2, etc...). Also, it is quite common to name the quantity obtained as a result of long productions in different ways in terms of convenience. The most common use of these types of custom-named quantities in Table2 with expressions in SI base units.

Quantity Name	**Unit name**	**Unit symbols**	**In other SI units**	**In SI base units**
Plane angle	radian	rad		$m.m^{-1}$
Solid angle	steradian	sr		$m^2.m^{-2}$
Frequency	Hertz	Hz		s^{-1}
Force, weight	Newton	N		$kg.m.s^{-2}$
Pressure, stress	Pascal	Pa	N/m^2	$kg.m^{-1}.s^{-2}$
Energy, work, heat	Joule	J	N.m	$kg.m^2.s^{-2}$
Power, radiant flux	Watt	W	J/s	$kg.m^2.s^{-3}$
Charge	Coulomb	C		A.s
Voltage	Volt	V	W/A	$kg.m^2.s^{-3}.A^{-1}$
Capacitance	Farad	F	C/V	$kg^{-1}.m^{-2}.s^4.A^2$
Resistance	Ohm	Ω	V/A	$kg.m^2.s^{-3}.A^{-2}$
Electrical conductance	Siemens	S	A/V	$kg^{-1}.m^{-2}.s^3.A^{-2}$
Magnetic flux	Weber	Wb	V.s	$kg.m^2.s^{-2}.A^{-1}$
Magnetic flux density	Tesla	T	Wb/m^2	$kg.s^{-2}.A^{-1}$
Inductance	Henry	H	Wb/A	$kg.m^2.s^{-2}.A^{-2}$
Radioactivity	Becquerel	Bq		s^{-1}
Luminous flux	lumen	lm	cd.sr	Cd
Illuminance	lux	lx	lm/m^2	$m^{-2}.cd$

Table 2 SI derived units with special names and symbols.

Dimensional Analysis

Dimensional analysis is a very powerful analysis method that is frequently used in basic sciences and engineering. It refers to the fact that the dimension of the right and left sides of equality written with the simplest definition must be the same as each other. If the two physical quantities are comparable to each other, they have the same unit. For example, the quantities with a magnitude of length [L] can be comparable even if they are expressed in different units of 2 cm to 3 inches.

On the other hand, 3 kg to 4 s cannot be comparable to each other because one is mass [M] and the other is time [T] (See Table 1).

Dimensional analysis can be used as a simple provisioning process, such as controlling the consistency of an equation given or derived, or for sophisticated purposes, such as investigating which physical quantities a physical quantity depends on. If we look at this with an example, let's try to figure out how long it is bound to the extent and how it will fall to the ground of a free object:

Let's write to the left side of our equation what this might depend on the right side of the object's fall time. These possible quantities can often be determined by rough observations. If we select the mass (m), the height (h), and the gravitational acceleration (g) where the object is dropped as possible quantities, we can assume an equation like the following,

$$t = C. m^{\alpha} . h^{\beta} . g^{\gamma} .$$

Here, C is a dimensionless mathematical constant that is likely to be included in the equation. The value of this can only be determined by an experiment, but it can be investigated by the analysis of the dimension of other unknowns, α, β, and γ. If we replace the dimensions of the left and right sides of the equation, we will achieve the following equality,

$$[T] = [M]^{\alpha}. [L]^{\beta}. \frac{[L]^{\gamma}}{[T]^{2\gamma}}.$$

The necessity for the dimensions of the left and right to be equal gives us a correlation about the exponent. First of all, it can be seen immediately that there should be $\alpha = 0$ because there is no mass neither on the left nor on the right. On the other hand, the right side of equality should neglect each other because there is no length on both sides. And from here $\beta + \gamma = 0$ is required. Finally, the fact that the time size on the left side is 1 and on the right side gives the equation $-2\gamma=1$, which includes $\gamma=-1/2$ and therefore $\beta=1/2$ from the previous equation. After all, we get the following equation for the object's fall time,

$$t = C. \sqrt{\frac{h}{g}}.$$

If we recall the $h = ½gt^2$ relation between the height of a free-falling object and the falling time from high school, we do not need to experiment to know that the constant C here should be $\sqrt{2}$. As can be seen, it is possible to draw out the outline of a physical correlation with only rough assumptions without the need for any physical background information. Of course, the success of the size analysis depends on the possible variables to be chosen at the beginning.

Example: Find the formula that gives the vibration frequency (f) of a guitar wire by analyzing the stress force (F) on the wire, the length of the wire (L), and the longitudinal density (ρ_{length}) of the wire (The height density means mass per length).

Uncertainty

No matter how precise any measuring instrument is, it is not possible to report the measured size with an infinite number of digits (hence an infinite precision), so all measurements must be understood and evaluated within a certain "range" of precision. This range is called the uncertainty of that measurement and is expressed by placing a "±" sign next to the measured value. As an example, it is understood that the length of a bar whose length is reported as 123.4 ± 0.3 mm has a value between 123.1 mm and 123.7 mm. Of course, it is possible to reduce uncertainty even more by using a more sensitive tool, but uncertainty can never be zero. Therefore, a measurement that does not contain uncertainty information cannot be evaluated scientifically.

While the uncertainties in the values read by some complex devices are written on the device, most devices show this information by the number of numbers they report. If uncertainty is not explicitly given in a device, it would be appropriate to accept that there is an uncertainty that is half the digit of the rightmost digit of the number read from the device. For example, a weight measured as 32.82 g in an electronic scale should be understood as 32.82 ± 0.005 g. (Since the digit of the last digit is one per cent, the uncertainty is written like half of it, i.e. 1/200 = 5/1000 = 0.005). As an example, the length measured by a ruler with millimetre lines is rounded to the nearest millimetre and the uncertainty is written as 0.5 mm.

Propagation of the uncertainty:

In physics, it is necessary to perform the measured quantities with each other. For example, if we want to measure the area of a plate with an indirect measurement, it will be necessary to measure the width and length separately and do with the multiplication of these lengths. As another example, when it is necessary to measure a long distance (and if the length of our meter is not enough for a single measurement), it will be necessary to take a few measurements and add up. In this case, it is necessary to know how these uncertainties will be reflected in the result when the values with different uncertainties are processed. The following rules apply to the four operations.

- Uncertainties are collected in summation and subtraction.
    - Summation:

$$(23,48 \pm 0,18) + (12,11 \pm 0,33) = 35,59 \pm 0,51$$

    - Subtraction:

$$\{(23,48\ \pm 0,18) + (12,11\ \pm 0,33) = 11,37\ \ \pm 0,51\}$$

- Percentage uncertainties are collected in multiplication and division. The result found refers to the percentage of uncertainty as a result of the operation.
    - Multiplication:

$$(23,48 \pm 0,18) \times (12,11 \pm 0,33) =$$

$$(23,48 \pm \%7,66) \times (12,11 \pm \%2,72) =$$

$$(23,48 \times 12,11) \pm (\%7,66 + \%2,72) = 238,34 \pm \%10,38 = 2384,34 \pm 29,51$$

    - Division:

$$\frac{23,48 \pm 1,8}{12,11 \pm 0,33} =$$

$$\frac{23,48 \pm \%7,66}{12,11 \pm \%2,72} =$$

$$\frac{23,48}{12,11} \pm (\%7,66 + 2,72) = 1,94 \pm \%10,38 = 1,94 \pm 0,20$$

In general, the Crank Three Times method can be used for the propagation of the uncertainty. In this method, the magnitudes that are processed are calculated with their nominal values, to “give the biggest result" and "to give the smallest result" and the uncertainty in the result is reported to contain these three results around the nominal value.

Example: sin(21,3±0,4)=?

Nominal value: sin(21,3)=0,363251

Biggest value: sin(21,7)=0,369746

Smallest value: sin(19,9)=0,340379

Result: sin(21,3±0,4)=0,363251(+(0,369746-363251)-(0,363251-0,340379))

sin(21,3±0,4)=0,363251(+0,006495-0,022872)

Significant figures and rounding

If we remember the instruments which uncertainties are not explicitly written, we have said that the observer should accept that there is an uncertainty that is half the digit of the rightmost digit of the reading number. When it comes to measuring, the concept of numbers now starts to mean different things for a mathematician and a physicist. For a mathematician, 1,200 and 1,2 mean the same, and when a physicist reads these numbers on two different measuring devices, the first number understands that there is a "hidden" uncertainty of ± 0,0005 and the second to ± 0,05. Therefore, it is necessary to avoid writing unnecessary digits or fractions when writing the measurement results because the last digit also determines your uncertainty. In another word, it is necessary to pay attention to whether each step is written. In this context, the numbers that can be accepted as meaningful in a number read in measurement are determined by the following rules:

- All digits different from zero are significant (123,45 → 5 significant figures).
- Zeros between non-zero digits are significant (10023,00405 → 10 significant figures).
- Leading zeros are never significant (00123,45 → 5 significant figures).
- Zeros are significant at the end of the decimal numbers, and not significant at the end of the integer numbers (12,300 → 5 significant figures, 12300 → 3 significant figures).

The only confusing thing here is the assumption that the number 12300 has 3 significant figures. Why this should be done will be explained in an example soon. But first, let's clarify the exceptional situation: If an observer reads this number exactly on the screen of the instrument and therefore wants to express that the trailing zeros are significant, it is best to switch to scientific notation and express this number as $1,2300\times10^4$ to avoid confusion[1].

The concept of significant figures greatly facilitates the transfer of uncertainties described above and relatively complicated when dealing with the measurement results. For this, it is enough to pay attention to the following rules.

- When the two numbers are multiplying or dividing, the result is rounded, which has at least significant figures, among those into entered the operation.

[1] To overcome this confusion, there are two notations such as putting an extra comma at the end of the number or drawing a line above the last significant figure, but none of them are widely used as accepted. It is safest to show it with scientific notation.

Examples:

Operation	Actual Result	Significant Result	Statement
8×8,00	64	60	Although the second term contains three significant figures, the first term has a single significant figure, so the result is rounding that it contains a single significant figure.
8/1,25	6,4	6	
8,0×8,0	64	64	The terms contain two significant figures, so the result has two significant figures.
8,6/2,0	4,3	4,3	
8,02×8,02	64,3204	64,3	The terms contain at least three significant figures, the result is rounded to include three significant figures.
12,250×21,3	260,925	261	

- In summation and subtraction, the result is rounded up to the last significant digit of the one with the highest order value among those who entered the operation.

Examples:

Operation	Actual Result	Significant Result	Statement
1+1,1	2,1	2	The first term has a significant figure in the first order so the result rounds to first order.
123+60	183	180	The second term has a significant figure in the second order so the result rounds to second order.
123,25+46,0+86,26	255,51	255,5	The second term has a significant figure in the decimal order so the result rounds to decimal order.
5,67-3	2,67	3	The second term has a significant figure in the first order so the result rounds to the first order.

One of the important points to be considered when dealing with significant figures is the need to exclude mathematical constants from significant figures evaluations. Because it expresses FULL certainty without a mathematical constant uncertainty (it can also be thought of as having an infinite number of significant figures). For example, if the $\frac{1}{2}mv^2$ equation is used in the kinetic energy calculation, the significant figures (and the processes between them) should be discussed through physical quantities m and v. It is WRONG to think of the ½ as a number with 1 significant figure. Either it should not be included in the evaluation at all, or it should be thought that it has an infinite number of significant figures such as 0.500000... (both approaches already give the same result). As a striking example of this, let's assume that the period of a pendulum is measured as 12.3 s. Since the frequency is defined by the formula 1/T, while expressing the frequency of this pendulum, the number 1 in the formula is not considered to have a significant figure because it is a mathematical constant. Therefore, the result is again rounded to 3 significant figures and expressed as 0.0813 Hz.

What to do with some special functions when dealing with significant figures is described below:

- When taking the power or root of the number, the result should have significant figures the same as the number.
- When the **ln(x)** or **log (x)** function is used, the result should maintain "decimals" as much as the significant figure of digits of x. Example: ln (8,3) = 2,1162555... The result is rounded to two meaningful numbers after the comma: 2,12
- In the case of $\mathbf{10^x}$, the result contains as many significant figures as the number of significant digits in the part of x after the comma. Example: $10^{4.3}$=19952,623…=2×10^4 should be rounded to a single meaningful number.
- For **the exponential function**, the number of significant figures is maintained. Example: $e^{5,32}$ = 204.
- For the **sine** function, the number of significant figures in the result is the sum of significant figures the angle and decimal order of the angel. Example: sin (34,21) = 0,562228.
- For the **cosine** and **tangent** functions, the number of significant figures is maintained. Example: cos (12,3) = 0,9770455… = 0,977.

Important note for rounding: One of the common mistakes are made when the rolling the numbers is starting from the rightmost digit. While this approach extends the process unnecessarily, it may even cause errors in some cases. The correct method is to determine the number of significant figures and then just round the last number.

Example: We want to round the number 25,874678 to 4 significant figures. If I start from the rightmost, this process gives me 25,88. However, this number is closer to 25,87. The correct method is, not to look after the 5th significant figure but to see the number as 25,874 from the beginning to round the number to 25,87.

Errors

The difference between the measured value and the actual value of a physical quantity is called error. Errors that can be encountered in experiments or observations are divided into two.

Systematic Errors:

Errors that deviate the result in the same direction with each measurement are called systematic errors. These can also be divided into three:

1. **Originated from experimental set-up:** These errors are mainly originated from calibration faults of testing apparatus. For example, non-precision scales, incorrect weight and non-calibrated thermometer etc.
2. **Implementation errors:** These types of errors are caused by the miss-use of instruments. Carrying out of the experiment thoroughly decreases these errors.
3. **Environmental errors:** It is necessary to pay attention to environmental effects such as temperature, humidity, wind etc. in the experiments.

Random Errors:

Random errors are errors that arise from the sensitivity of the measuring instruments or the border of the sensory organs of the observer, the direction of which cannot be estimated and are found in every measurement. Since they are purely coincidental, the probability of these errors to be positive or negative is equal. Even if their causes are known, they cannot be eliminated. The most effective way to deal with these errors is to take many measurements and use the statistical methods mentioned below[2].

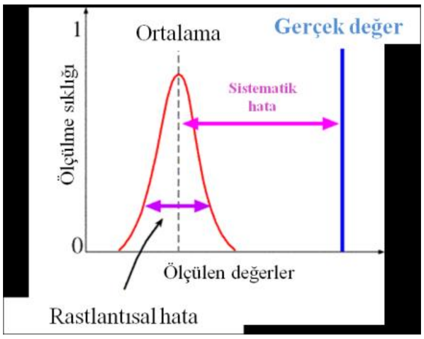


Figure 2

The arithmetic mean and standard deviation

Random errors, by definition, divert the measurement results with equal probabilities above and below the true value. Therefore, it is the most reasonable way to take the arithmetic average of the measured values while searching for the real value[3]. If we represent the magnitudes we find as a result of n measurements with $x_1, x_2, \ldots, x_n$, the arithmetic mean of this data set is represented by the symbol $\bar{x}$ and is calculated by the formula below,

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Let's continue with an example. The two observers, who performed the same experiment with different methods, took 10 measurements to reduce random errors and obtained the numerical data shown below:

Number of observations	1st observer	2nd observer
1	15,08	14,7
2	15,02	15,2
3	14,91	15,1
4	14,86	14,9
5	15,06	15,0
6	14,77	14,2
7	15,22	15,7
8	14,90	15,3
9	15,12	14,8
10	15,06	15,1
Mean	**15,00**	**15,0**

Table 3

[2] These statistical methods are useless in dealing with systematic errors, they only ensure that random errors are kept under control.

[3] If we take the systematic error as zero, it can be proved mathematically that the arithmetic mean is the "most likely" approximation for the true value, but we do not take this proof here. For further reading: Basic Data Analysis for Experiments in the Physical Sciences, Erhan Gülmez.

Since these two observers used different methods, probably because of the sensitivity of their devices, the first one could express two significant numbers after the comma, while the second could only write a significant number. If you calculate the average of both data sets, we find 15.00 for the first and 15.0 for the second. Although the average of the two measurements is numerically the same, there is a clear difference between measurement uncertainties, and we need a definition to express this difference in some way when examining random errors. (This definition is also a size of the "width" of the red line in Figure 2.) This definition is called the ***standard deviation***.

The standard deviation is an indicator of how much the values in a data set deviate from their averages "on average". The standard deviation is indicated by the symbol σ and is calculated using the formula below[4].

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

Continuing our example, the first observer calculated the average as 15,00. To calculate the standard deviation, it must subtract each squared value from this average and calculate its square root after dividing their sum by the number of data. This process is shown in Table 4 and below.

x	$x - \bar{x}$	$(x - \bar{x})^2$
15,08	0,08	0,0064
15,02	0,02	0,0004
14,91	-0,09	0,0081
14,86	-0,14	0,0196
15,06	0,06	0,0036
14,77	-0,23	0,0529
15,22	0,22	0,0484
14,9	-0,10	0,01
15,12	0,12	0,0144
15,06	0,06	0,0036
SUM	**0,00**	**0,1674**

Table 4

$$\sigma = \sqrt{\frac{0{,}1674}{10}} = 0{,}13$$

As can be seen from the table, the difference of values from the average cannot be used as a measure of deviation alone, since their sum always gives zero (it can be easily demonstrated mathematically that this should always be so[5]). Therefore, the reason for squaring the differences becomes clear[6]. Since the squares of the differences from the average will have the square of the dimension of the measured quantity, the square root of the final must, of course, be taken for the result to be the same dimension (and comparable) with the value.

[4] The definition we give here is called the sample standard deviation. There are different standard deviation definitions for different purposes. For further information: https://en.wikipedia.org/wiki/Standard_deviation

[5] $\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$. The first term here is the sum of the values, that is, it can be written as $n.x$, which is the product of the average by the number of values. Since the average in the second term is already calculated, it goes out of the sum, and $n.x$ is again obtained. So, the difference between the two is zero.

[6] It is conceivable to use the absolute value to get rid of zero, but the derivative of the absolute value is not continuous like a square. Another advantage of taking a square is that it "punishes" large differences by making it larger.

If we apply the above operations to the results of the second observer (do this calculation yourself and confirm the result) we see that the standard deviation is 0,38. Therefore, the difference in precision that we can perceive by looking at the raw data of the two observers shows itself quantitatively in standard deviations.

Where a large number of measurements are taken, as in this example, the result is usually reported in the format $x \pm \sigma$. In this case, while the first observer writes 15,00 ± 0,13, the second observer should round the standard deviation, as in his data, to a significant figure after the comma and write it as 15,0 ± 0,4.

Graph drawing

In positive sciences, experiments are carried out by monitoring and recording how one physical quantity is controlled in a controlled manner and affects another quantity. These records are usually taken in a tabular format, as will be shown in the example below. However, at the end of the experiment, looking at the numbers in these tables and trying to discover the relationship between them is not a suitable process for the human brain. People can comprehend relations more efficiently by visual means (which directly address the most important of the five senses) rather than abstract numeral symbols. Therefore, it is preferred to visualize the data in some way. This visualization process is called graphic drawing.

A graphic is constructed by drawing two number lines (axis) selected perpendicular to each other. One of these axes represents changing quantities in the experiment, and the other represents measuring quantities. The data represents the points of the numerical values corresponding to each other to the axes. Then the best curve is proposed to represent the distribution of these points. The parameters that optimize the selected curve are calculated, and the process ends withdrawing the curve.

Graphics are drawn on special papers called graph paper. An example blank graph paper is shown below.

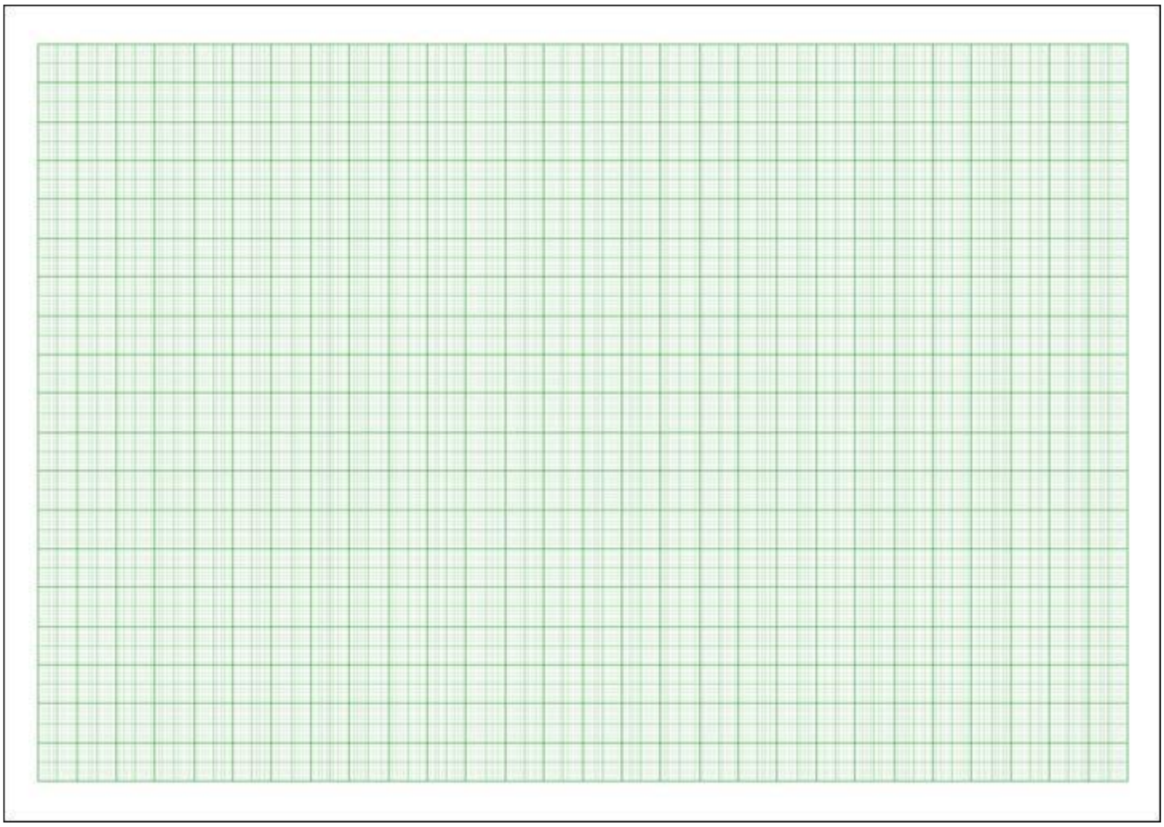


Figure 3 Graph paper

Let's show the rules required to draw a smooth chart step by step on an example. Let's assume that the position (x) of an object travelling at a constant speed in a mechanical experiment is measured against time (t). The result of this measurement is shown in the table below.

t (s)	x (m)
0	0,8
1	1,6
2	3,5
3	6,0
4	7,8
5	11,2
6	12,0

Table 5 Table of an example measurement in which the position versus time is recorded.

We want to analyse this data on a graph. The first step in drawing graphics is to identify the axes.

1. **Identifying axes:** By selecting the origin near the lower-left corner of the graph paper, axes are drawn from here to the right and up. One of them is identified time or t and the other position or x. The name can be written as a word or symbol. In any case, the unit of these quantities must be written in parentheses. A graphic whose axes are not determined

or written units has no scientific value.

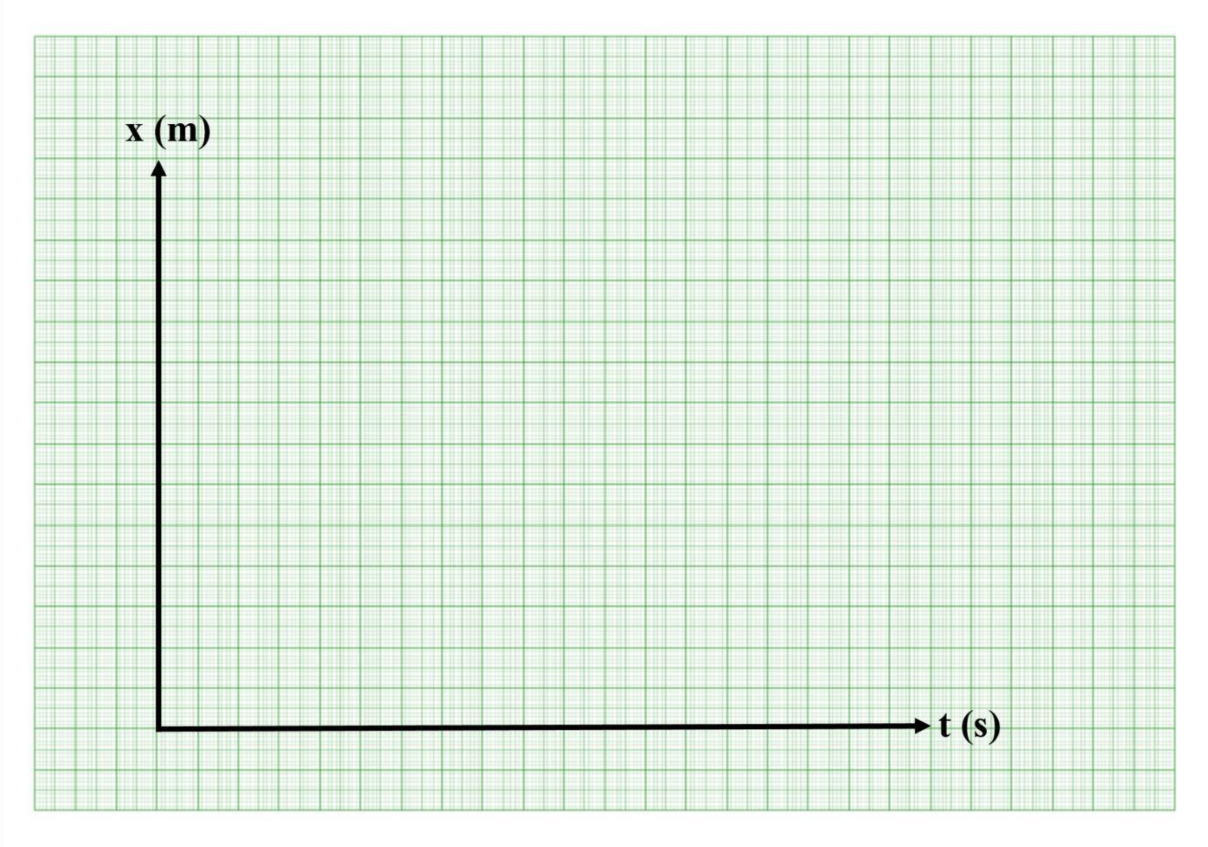


Figure 4 Drawing and determining the origin and axes. Axis symbols and units should never be forgotten.

2. **Scaling:** After drawing the axes, naming them, and writing their units, what needs to be done is to scale them appropriately. What is meant by this is above all the determination of how many points correspond to the axis from the quantity represented on that axis. A well-scaled axis is scaled to include the entire set of data given, without overflowing. The best way to do this is to look at the range of data represented on that axis and compare this difference with the length of the axis.

   Returning to our example, we see that t values are between 0 and 6. On the other hand, since the axis we set for t in our graph is a bit longer than 18 cm, it is appropriate to take every 3 cm on the axis as 1 s. The x values we represent on the vertical axis range from 0 to 12, and the length of the axis is around 13-14 cm. Therefore, 1 m is placed in every cm.

   After this analysis is done, some numerical values are written on the axes that can help in drawing. The point to be considered here is to choose a range that is frequent enough to provide ease of drawing, but the axis is as crowded. In the graphic below, a value of 3 cm is written on both axes. For such a graph paper, it can be said that 3 or 4 cm is the ideal range. The important thing here is to maintain the balance between ease of drawing and the unnecessary crowd of numbers.

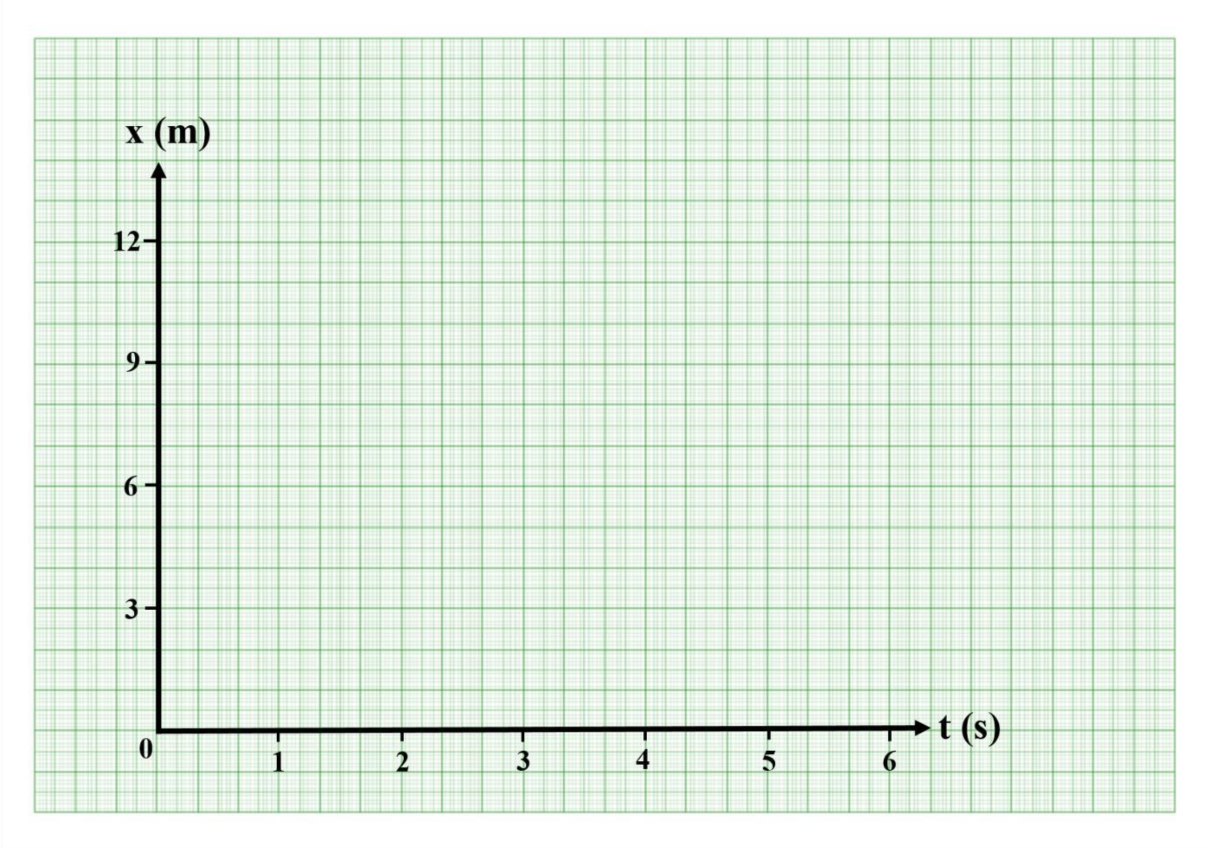


Figure 5 Scaled graphic.

3. **Placing the points:** After scaling the axes, points can be placed. While doing this, the lines of the graph paper are used. Therefore, there is no need to draw auxiliary lines indicating where the points correspond to the axes or to write numbers indicating the values of the points on the axes. The person who will read the graphic can find this information through the lines of the paper if he wishes. A scientific notation should be complete but simple.
For the points, a size that can be easily selected by the eye should be selected even if a

curve is passed over it.

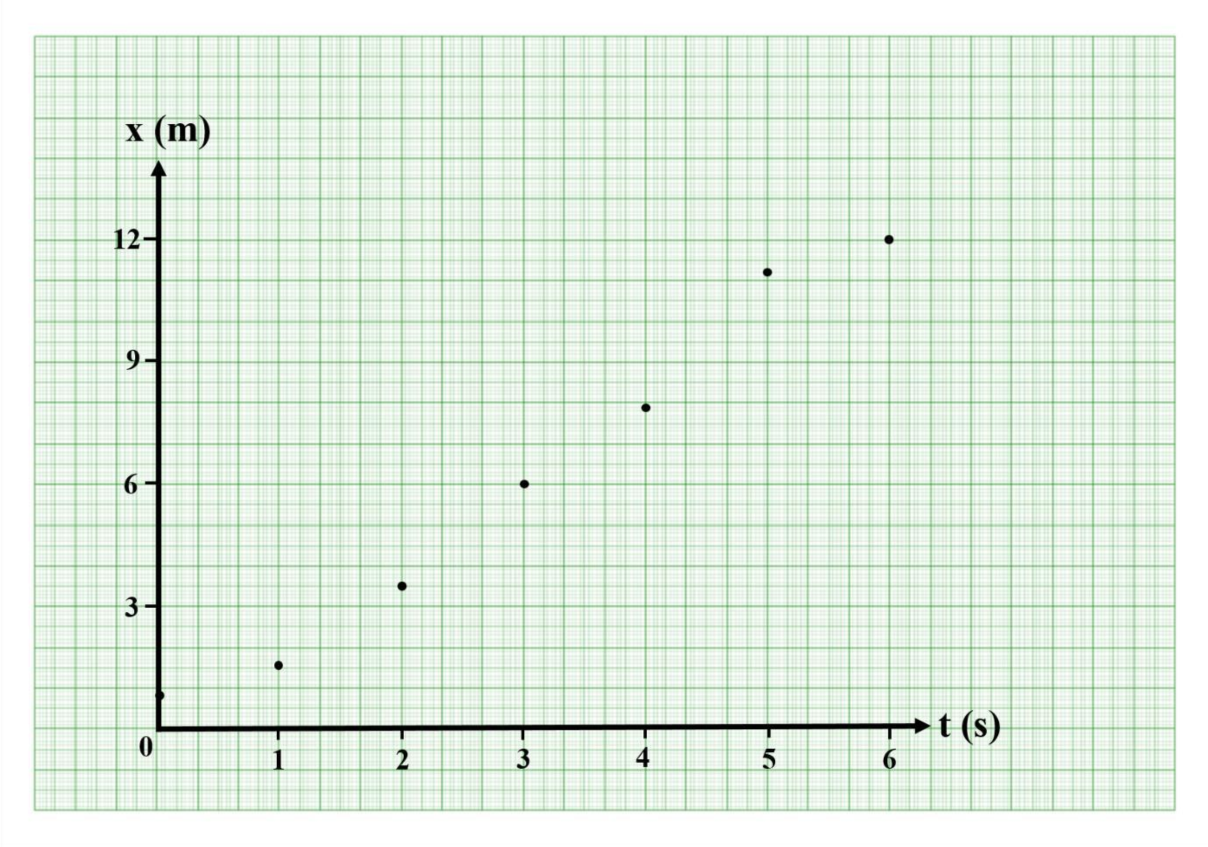


Figure 6

4. **Fitting:** Placing the dots does not complete the graphic drawing process. The main thing is to show the relation represented by these points with a continuous curve on the graphic. Let's examine this process under three subheadings.
   i. **Proposing Equations:** Relationships between physical quantities are derived from physical theories. We said at the beginning that our example is uniform motion. The kinematic formula that gives the relation between position and time in the uniform motion is as follows.

$$x = x_0 + v_0 t$$

   This expression is a linear equation according to the variables *t* and *x*. x_0 and v_0 are the parameters that define this line[7].
   Whichever theory is desired to be tested in an experiment; it is necessary to plot the curve represented by the relation derived from that theory together with the data points. Thus, how far the theory (line) fits the experiment (points) is visualized.
   After determining with which relation, the data will be represented, the next step is to find the parameters that make our equation "most suitable" for these data.
   ii. **Fitting:** The least-squares method is one of the most widely used methods in the process of determining the parameters of the selected curve. In this method, the "best suitability" criterion is defined as the smallest sum of squares of the differences between the dependent variables of the data points and the dependent variables of the curve. When this definition is expressed mathematically, it is

[7] These mathematically correspond to the point intersects the x-axis respectively and slope where the line, while physically corresponding to the initial position and velocity of the object.

possible to find equations that express the parameters of the selected curve in terms of data (i.e. coordinates of the points)[8].

In this context, if a line is selected as $y = mx + n$ as a correlation, m and n parameters of this line are calculated in terms of data points with the following expressions[9].

$$m = \frac{k\sum_{i=1}^{k} x_i y_i - \sum_{i=1}^{k} x_i \sum_{i=1}^{k} y_i}{k\sum_{i=1}^{k} x_i^2 - \left(\sum_{i=1}^{k} x_i\right)^2}$$

$$n = \frac{k\sum_{i=1}^{k} x_i^2 \sum_{i=1}^{k} y_i - \sum_{i=1}^{k} x_i y_i \sum_{i=1}^{k} x_i}{k\sum_{i=1}^{k} x_i^2 - \left(\sum_{i=1}^{k} x_i\right)^2}$$

In this formula (x_i, y_i) pairs are the data points and k is the total number of data points. In our example, let's rename the variables and parameters as below to use these formulas.

$$t \to x$$
$$x \to y$$
$$v_0 \to m$$
$$x_0 \to n$$

Let's calculate the four sums that should be used in formulas m and n:

$$\sum_{i=1}^{k} x_i = 0 + 1 + 2 + 3 + 4 + 5 + 6 = 21$$

$$\sum_{i=1}^{k} y_i = 0{,}8 + 1{,}6 + 3{,}5 + 6{,}0 + 7{,}8 + 11{,}2 + 12{,}0 = 42{,}9$$

$$\sum_{i=1}^{k} x_i^2 = 0^2 + 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = 91$$

$$\sum_{i=1}^{k} x_i y_i = 0 \times 0{,}8 + 1 \times 1{,}6 + 2 \times 3{,}5 + 3 \times 6{,}0 + 4 \times 7{,}8 + 5 \times 11{,}2 \ldots + 6 \times 12{,}0 = 185{,}8$$

If we write these sums in the formula of m and n, we find $m = 2{,}02$ and $n = 0{,}01$. If we go back to the original names of variables and parameters, we see that the line that best fits our data has the equation $x = 0{,}01 + 2{,}03t$.

iii. **Curve sketching:** The curve whose parameters have been determined as the last stage can now be drawn on the graph paper[10]. As in our example, if this curve is a straight line, the drawing is relatively easy: Two points that the line passes over are determined and these points are marked on the graph paper (so as not to interfere with the data points), and then the curve is drawn with the ruler. The

[8] For further information about the least-square method: https://www.youtube.com/watch?t=1&v=T48F7_e5sfM

[9] Although the m and n formulas we give here seem to work only for linear relations, they also allow a much wider family to be examined if used wisely. In a situation where there is a general correlation between x and y values such as $y = A.x^{\alpha}$, taking the logarithm of the two sides gives us the $\log(y) = \log(A) + \alpha.\log(x)$ equation, which is nothing more than a linear equation. Therefore, if the logarithms of the data are examined with the same method instead of the data themselves, the linear fit formulas give us the parameters A and α.

[10] It may not be easy to draw the curves other than the straight line, in this case, as many points as can be determined to outline the curve.

equation of this line is indicated on the graph paper.

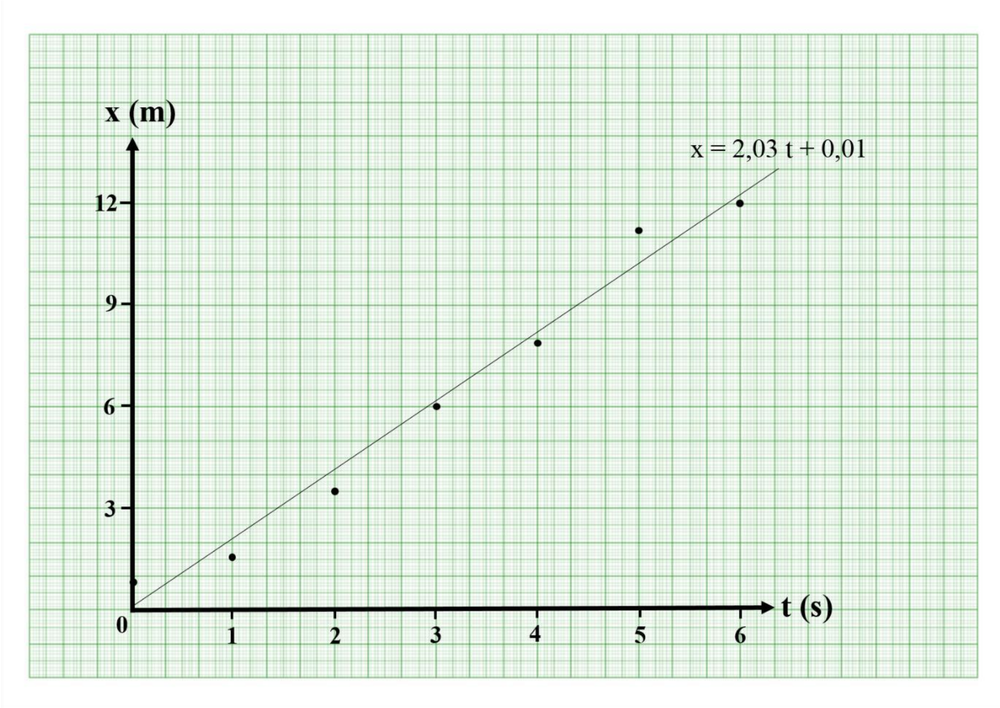


Figure 7

References

Hacettepe University FİÖ 213 Physic Laboratory Booklet

Basic Data Analysis for Experiments in the Physical Sciences, Erhan Gülmez, Bogazici University Publications, July 1997

https://en.wikipedia.org/wiki/Significant_figures

https://en.wikipedia.org/wiki/Significance_arithmetic

http://www.av8n.com/physics/uncertainty.htm#sec-crank3

http://fabice.com/misc/significant_figures.html

https://en.wikipedia.org/wiki/Standard_deviation

https://www.youtube.com/watch?v=BnYQAJIhTf0 **(Linear fitting)**

https://www.youtube.com/watch?v=T48F7_e5sfM **(The least-square Method)**

https://www.youtube.com/watch?v=N8WekH6zLRM **(Fitting with MS Excel)**

Appendix: The least-square method

Suppose that the equation of the selected curve to fit a data set has a general expression such as $y = f(x)$. Let's assume that f(x) contain various parameters such as a, b, c, d, …

The least-squares method defines the condition that this selected curve should be “best suited” to the data set as follows: The sum of the squares of the differences between the dependent variables of the data points and the dependent variables of the curve should be the smallest.

In this context, we have data and we want to fit the $y = f(x)$ curve to these points. If we show the independent variables with x and dependent variables with y, the difference mentioned in the definition for a selected (x_i, y_i) data point can be written as $y_i - f(x_i)$. Let's find these differences for all data points, add the squares and call this S.

$$S = \sum_{i=1}^{k} \left(y_i - f(x_i)\right)^2$$

Since x_i and y_i are numbers that we know from the experiment, the quantities to which S is linked are the parameters in the $f(x)$. As it is stated in the definition, it is necessary to take the partial derivative of S according to each parameter and to equal zero to find the parameters that will make the sum minimum.

$$\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0, \frac{\partial S}{\partial c} = 0, \ldots$$

As a result, it is possible to generate as many equations as shown above, and the solution of these equations gives us the numerical values of the parameters sought.

This is the theory of the least-squares method. From here on, let's assume that it expresses a line equation $f(x) = mx + n$, as an example. In this case, we can write the S sum as follows.

$$S = \sum_{i=1}^{k} (y_i - mx_i - n)^2$$

Now if we take the partial derivatives according to the m and n parameters, whose values we seek, we find the following expressions.

$$\frac{dS}{dm} = \sum_{i=1}^{k} -2(y_i - mx_i - n)x_i$$

$$\frac{dS}{dn} = \sum_{i=1}^{k} -(y_i - mx_i - n)$$

Since both equations will be equal to zero, they can be written with simplifications as follows.

$$\sum_{i=1}^{k} (y_i - mx_i - n)x_i = 0$$

$$\sum_{i=1}^{k} (y_i - mx_i - n) = 0$$

When we distribute the total signs inside, it is possible to edit our equations as follows.

$$m\sum_{i=1}^{k} x_i^2 + n\sum_{i=1}^{k} x_i - \sum_{i=1}^{k} x_i y_i = 0$$

$$m\sum_{i=1}^{k} x_i + n\sum_{i=1}^{k} 1 - \sum_{i=1}^{k} y_i = 0$$

Therefore, we have obtained a linear system of equations consisting of two equations for the unknowns *m* and *n*. When this system of equations is solved with the desired method (complete this solution) the following equations for m and n can be obtained on page 15.

$$m = \frac{k\sum_{i=1}^{k} x_i y_i - \sum_{i=1}^{k} x_i \sum_{i=1}^{k} y_i}{k\sum_{i=1}^{k} x_i^2 - \left(\sum_{i=1}^{k} x_i\right)^2}$$

$$n = \frac{k\sum_{i=1}^{k} x_i^2 \sum_{i=1}^{k} y_i - \sum_{i=1}^{k} x_i y_i \sum_{i=1}^{k} x_i}{k\sum_{i=1}^{k} x_i^2 - \left(\sum_{i=1}^{k} x_i\right)^2}$$

When it is desired to derive formulas for the parameters of the curves other than the straight line, go to the step where the sum of S is written, replace the expression $f\ (x)$ here and take the partial derivatives to derive the equations, and then these equations must be solved algebraically.