Contents:

INTRODUCTION

The object of performing the physical experiment is usually to obtain a set of numerical results. Gathering data during the measurements is only the first step on the way to give the final numerical result of the experiment. After that there are processes of arithmetical calculation that involve application of physical formulas as well as the averaging or smoothing the measured values.
The treatment of the experimental data is not ended when we have obtained desired numerical result. An important part of the job is to determine the degree of uncertainty of the proposed result.

When one measures a physical quantity he does nor expect the obtained value to be exactly true value. It is crucial to indicate how close the result is likely to be to the true value. It is done by including an estimate of the error made with the reported result.

Every physical quantity is subject to error in its determination. Errors can be generally defined as the difference between the experimentally obtained and the true value. Usually we do not know this "true" value. We may only know approximately what it should be from earlier experiment or from theoretical predictions. This can inform us whether our result is of the right order of magnitude, but we should determine from the data and from the experimental conditions how much confidence we can have in obtained results.
Once the result of the experiment has been reported it becomes a "public property" and can be used by others in different ways.
The reported value of the physical quantity, when not accompanied by a statement of its uncertainty, can be of small value.

Assessment of the uncertainty involves some knowledge of the accuracy and precision of the experimental method and instruments used, and a study of how the uncertainties in the gathered experimental data affect the final result.
These issues are not of peripheral interests in the experiments. On the contrary they are related to the purpose of the experiment, used methodology and significance of the obtained results.

PRECISION AND ACCURACY

For any experiment both precision and accuracy should be simultaneously considered.
The "precision" of an experiment tells us how precise the result has been determined without giving the degree of agreement with the true value.

Absolute precision $\Delta\mathbf{x}$ indicates the magnitude of uncertainty in the result and is given in the same units as the result. It is defined by the difference between the measured value - $\boldsymbol{x}$- and the true value - $\boldsymbol{X_0}$ - Eq.1.:

$$\Delta x = X_0 - x \qquad (1)$$

Relative precision $\Delta\varepsilon$ indicates the uncertainty as a fraction of the value of the result (mostly in %). It is defined by the division of the absolute precision by true value - Eq.2:

$$\Delta\varepsilon = \frac{\Delta x}{X_0} \qquad (2)$$

In many cases the relative precision, because of its relative relation with the true value gives more adequate information about the certainty of the measurement than the absolute precision.

Usually the precision with which the result is determined is referred to the uncertainty in the result. Term "accuracy" of the experiment is a measure of how close to the true value is the obtained result.

SOURCES OF EXPERIMENTAL ERRORS

Illegitimate errors.

Errors that come directly from the mistakes made by the observer during the measurements or calculations are classified **as illegitimate errors.**
These blunders are usually easy to recognise because they appear as unreasonably data points or as the results far away from expected values. Once found they can be generally corrected and avoided in the future.

Systematic errors.

Errors that come either from the characteristics of the used instruments or of the experimental procedure itself are called **systematic errors**.
Examples of systematic errors are : calibration error of the instrument, uncompensated instrumental drift, parallax, scale shift, background, observer bias.
Such errors can be sometimes difficult to identify . The best way to recognise them is to compare results of several entirely independent experimental methods. Such results are different from the other with reproducible discrepancies.
Systematic errors can be eliminated by improving the calibration precision, sufficient attention to controls and taking readouts or even by corrections introduced during the calculations themselves.
The ability of recognition, compensation and control the systematic errors results in improvement of the accuracy of an experiment.

Random errors.

Errors that represent the fluctuations in results of series of observations are called **random errors**.
Independent readings of the same quantity differ by small random amounts. It is due to the observer inability to discriminate between readings differing by small amount and inability to increase the precision of the instrumental adjustments or to unpredictable fluctuations in the environmental conditions.
When the random errors come from instrumental uncertainties they can be reduced by using more accurate measuring equipment. If the result come from statistical fluctuations associated with counting finite numbers of events they can be reduced by increasing the number of measurements

The **precision** of the result, concerned as its reproducibility, is referred to the random errors, while the **accuracy** is an expression of the total uncertainty including that due to the systematic errors.
It is possible to have high precision and low accuracy when the significant systematic error and small random error occur at the same time. Because of their origins systematic errors may be orders of magnitude greater than the random errors, so the avoidance of systematic errors is very important during the planning stage of the experiment and taking the measurements.

CALCULATION OF EXPERIMENTAL ERRORS

Directly measured quantities - single measurements

Number of physical quantities e.g.: weights, volumes, temperatures, electric current, are established by direct measurements. In many experiments these directly obtained values are involved in later calculations of indirectly determined quantities. the propagation of different types of errors will be discussed later here.

Now, lets consider the precision of the directly obtained measurements. The establishing of the random error of such result needs a number of measurements of the same quantity (the greater number the better estimation). In various cases there is need to determine the precision of a single measurement of the quantity - ***x.***
The absolute precision (or **absolute experimental error**) of the single measurement - Δx- is defined in Eq. 1. Since we do not know the true value - X_0 - the special estimation procedure based on the precision of the measuring equipment used should be involved. It depends on the functioning principle of the instrument.
For the readouts taken by the comparison of the object parameter (e.g. border, mark) with the instrument's scale (e.g. Vernier calliper, thermometer, stopper) the absolute experimental error of the single measurement is given by the value of the half of the smallest division of the used scale.
The **relative experimental error** - $\Delta\varepsilon$ - (Eq.2) in this case is given by the ratio of the absolute error to the measured value.
For the multi-range indicating instruments (e.g. voltmeters, ammeters) the absolute experimental error can be determined using the currently used range - ***R-*** and device's precision class - **C -** from the Eq. 3.:

$$\Delta x = R \cdot C \cdot 0.01 \qquad (3)$$

EXAMPLE 1 .

To estimate the absolute experimental error of single measurement of the temperature made by use of the mercury thermometer with the 0.5^0 divisions on the scale.

Solution:

$$\Delta x = \frac{1}{2} \cdot 0{,}5^0 = 0{,}25^0$$

EXAMPLE 2.

To estimate the relative (percentage) error of single measurement of the electric current made by multirange indicating ammeter with the precision class 1.5 , using the 50 mA range . The recorded current value was 30 mA.

Solution:

1. first we calculate the absolute error: $\Delta x = 50mA \cdot 1{,}5 \cdot 0{,}01 = 0{,}75mA$

2. then we calculate the relative error : $\Delta\varepsilon = \frac{0{,}75}{30} = 0{,}025$

$\Delta\varepsilon(\%) = 0{,}025 \;\; 100\% = 2{,}5\%$

Clearly, the procedures described above give the greatest possible precision of single directly made measurement. Due to the environmental disturbances in most cases this "optimistic" error values should be modified (increased) by the observer himself, according to the real possibilities of precise readouts.

CALCULATION OF RANDOM ERRORS

The Gaussian normal distribution

Let's consider a large number of measurements of physical quantity - X- with values - x_i (i=1, 2, 3....N) . Each measurement is subject to random error Δx_i. For simplicity of explanation we shall preliminary assume that the true value - X_0 - is known, so the errors Δx_i are also known.

The frequency of occurrence of errors of size Δx_i - $n(\Delta x_i)$ can be represented by means of a bar graph in which the argument scale (error) is divided into ranges of equal width and the height of each bar represents the frequency (number of particular measurements) of occurrence of measurement yielding errors that fall within the respective range.- see Fig.1.

Typical bar graph shows irregularities due to the statistical fluctuations and finite number of measurements. From the bar graph of this type we can make rough predictions concerning the probability that certain measurement will yield an error of a given value.

For more reliable predictions we should increase the number of measurements and simultaneously decrease the width - w - of the range. Therefore the statistical fluctuations will become smaller in relation to the height of the bar, so we can even draw the smooth curve through the tops of the bars and consider it to be a representation of the error probability function - $P(\Delta x)$.

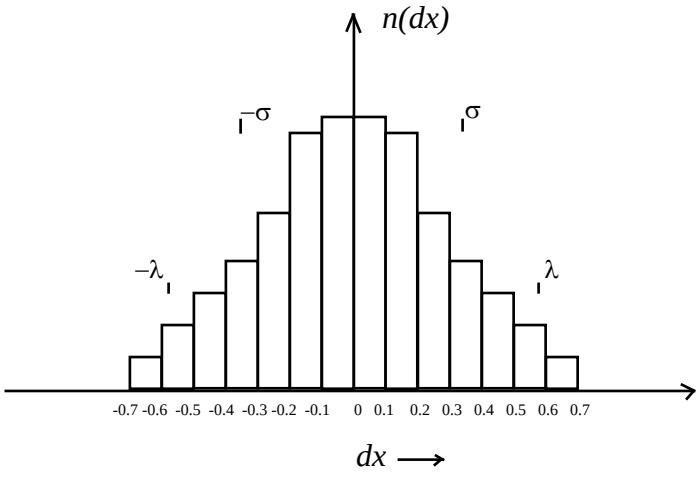

Fig.1. A typical distribution of errors.

Such function should be normalised to meet the condition - Eq.4.:

$$\int_{-\infty}^{\infty} P(\Delta x)\, d(\Delta x) = 1 \qquad (4)$$

The probability that a single measurement is subject to the error value from the range between ***Δx*** and ***Δx+ d (Δx)*** is equal to the ***P(Δx)d(Δx)***

The probability function introduced in this way is approximate because of a finite number of measurements. In many cases it can be assumed that the probability function is represented by so called **normal error probability function** (Gaussian normal distribution).- Eq.5.:

$$P(\Delta x) = \frac{1}{\delta \cdot \sqrt{2\,\pi}} \cdot \exp\left(\frac{-(\Delta x)^2}{2\sigma^2}\right) \qquad (5)$$

where:

δ is a parameter called the **standard deviation** .

It can be estimated from the set of error values - Eq.6.:

$$\delta^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (\Delta x_i)^2 \qquad (6)$$

One should notice the following peculiarities of this type of the distribution curve (see Fig.1.):

1. Deviations of the same absolute value are equally probable (the curve is symmetrical).
2. Small errors are more probable than great (the curve maximum is at about "0")
3. The greater δ value , the curve is more flat and it displays higher spread of measurements.

The normal error probability function is derived from several assumptions:
- that an error in measurement is compounded of the large number of unpredictable independent variable contributions (they do not conform to any particular probability function themselves);
- that on the average all of them are small in comparison with the error itself.

The real factors that determine the distribution are obviously unknown in detail. Thus, it is a common practice to assume that the normal error probability function may be applied without additional requirements. However, we should point out that in many cases with a great number of measurements the distribution curve may differs from that in Fig.1. (it may be shifted or flattened or multi-peaked etc.), so particular probability function should be derived.

In real measurements we seldom know the true value of measured quantity and so the errors - Δx - by which they deviate from true value . Instead of this kind of error we can only deal with the error from the "most probable value" of the quantity- x -. In terms of the Gaussian distribution the arithmetic mean value - x - is the best estimation of the "most probable value " of the quantity - x - and subsequently the errors - Δx - are estimated with - $\bar{x}$ - instead of the true value. So, the probability function would be - Eq.(7).:

$$P(x-\bar{x}) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp-\left(\frac{(x-\bar{x})^2}{2\cdot\sigma^2}\right) \qquad (7.a)$$

$$\sigma = \left(\frac{1}{N}\cdot\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^{\frac{1}{2}} \qquad (7.b)$$

$$\bar{x} = \frac{1}{N}\cdot\sum_{i=1}^{N}x_i \qquad (7.c)$$

The standard deviation of the mean - σ - and the arithmetic mean - $\bar{\mathbf{x}}$ - are called : " parameters of the normal Gaussian distribution".

In particular cases the standard deviation can be used as a measure of the uncertainty of the result. For large enough number of data (at least 30, preferably 100) with the ideal Gaussian distribution, the 68% of all the observations fall within $\pm\sigma$, and 95% of all the observations fall within $\pm 1.96\ \sigma$

In real most experiments are based on less than 10 measurements of the quantity and classical probability methods are inadequate for the treatment of them. Having small amount of observations with certain amount of spread that include the contribution of the random errors we wish to have the true value and the parameters of the hypothetical infinite population of measurements.
There are some techniques to deal with this problem, however all of them give results more or less approximate and limits of their usability we should always keep in mind.
There are four main ways in common use of expressing the precision of the data in terms of scattering of the individual values - $\mathbf{x_i}$ - about the mean.:

The **range** - ***R*** - "amplitude " of the dispersion - the difference between largest and smallest obtained value of - x_i - Eq.8.:

$$R = x_{\max} - x_{\min} \qquad (8)$$

The range is obviously a crude measure of precision ("the smaller - R - the more precise measurements") although it can be used for quick estimate of more sophisticated measures (it is discussed elsewhere here).

The **average deviation - *D* -** the average difference between the experimental values and their mean without regard to the sign-Eq.9.:

$$D = \frac{1}{N} \cdot \sum_{i=1}^{N} \left| x_i - \bar{x} \right| \qquad (9)$$

It is widely used for its simplicity however, it could be useful only for qualitative comparisons ("the smaller the better") . It is very sensitive to the size of sample (number of measurements).

The **quartile deviation - *r* -** (the " probable error") the value of the deviation from the mean that fulfil the condition that probability of errors smaller than - ***r*** - is equal to that for greater than - ***r* -** and it is 50%. It cannot be calculated directly unless there is a large number of observations (look for further details).

The **variance** - S^2 **-** (the mean quadratic error of single measurement) is defined by - Eq.10.:

$$S^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} \left(x_i - \bar{x} \right)^2 \qquad (10)$$

The divisor - (N-1) is the number of independent data on which the calculation (10) is based and it is called the "**number of degrees of freedom**".
Variance is unbiased by sample size can be directly useful in further probability calculations because it is in fact the standard deviation (its square) for the finite set of N data points. The square root of the variance is the best estimate of the true standard deviation - σ - of the hypothetical infinite population which can be made from a finite data set.

One can obtain an approximate relation between the variance - ***S* -** (10) and the average deviation - ***D*** - Eq. 11, and between the variance and the quartile deviation - ***r*** - Eq. 12.:

$$D = 0{,}7979 \cdot S \qquad (11)$$

$$r = 0{,}6745 \cdot S \qquad (12)$$

The Student method.

With the variance - ***S*** - value one can calculate so called ***P%*** **percent confidence interval,** i.e. the interval about the mean **- $\bar{x}$** - that ***P%*** of all ***N*** measurements would fall within it . For example in the ideal experiment (i.e. for great number of samples) the 95% confidence interval can be calculated with the variance as Eq.(13):

$$\bar{x} \pm \frac{S}{\sqrt{N}} \qquad (13)$$

For small samples (commonly in use) the random experimental error will introduce an additional uncertainty, which should be taken into account.
The solution of this problem is given in so called "Student" statistical distribution.
For each value of N there is an appropriate **correction factor** designated as - $\mathbf{t_p}$ **-** (in the table 1. there are given values for various degrees of freedom: ***N-1***).
The resulting confidence limit ($\pm \lambda_{\mathbf{p}}$) is given by - Eq.14.:

$$\lambda(P\%) \;=\; t_P \cdot \frac{S}{\sqrt{N}} \qquad (14)$$

where: $\mathbf{t_P}$ - is the correction factor (see table 1).

It means that : for ***N*** measurements of - ***x*** - with the distribution described by the variance - ***S*** - (Eq.10) the interval about the mean - $\bar{x}$ - (Eq.7.c) with the - ***P%*** - confidence is given by the Eq. 15.:

$$\bar{x} \pm t_P \cdot \frac{S}{\sqrt{N}} \qquad (15)$$

or, the probability that the deviation of measurement - $\mathbf{x_i}$ - from the mean value $\bar{\mathbf{x}}$ is within the limits of $\lambda = t_p \frac{S}{\sqrt{N}}$ is equal to ***P%.***

This is a thoroughly objective and unambiguous method of expressing the precision of a mean value.

Table 2. Critical values of correction factor t_p

degrees of freedom N-1	t_p for ***P%*** confidence					
	50%	80%	90%	95%	99%	99.9%
1	1.0	3.08	6.31	12.7	63.7	637.0
2	0.816	1.89	2.92	4.3	9.92	31.6
3	0.765	1.64	2.35	3.18	5.84	12.9
4	0.741	1.53	2.13	2.78	4.60	8.61
5	0.727	1.48	2.01	2.57	4.03	6.86
6	0.718	1.44	1.94	2.45	3.71	5.96
7	0.711	1.42	1.89	2.36	3.5	5.4
8	0.706	1.40	1.86	2.31	3.36	5.04
9	0.703	1.38	1.83	2.26	3.25	4.78
10	0.700	1.37	1.81	2.23	3.17	4.59
15	0.691	1.34	1.75	2.13	2.95	4.07

20	0.687	1.32	1.72	2.09	2.85	3.85

Estimate based on the range - *R*

For small samples (2 < N < 10) a reasonably accurate estimate of variance can be obtain by means of the range - **R** - (Eq.8). This technique has been developed out of the methodology known as the "quality control".
The approximate but very useful estimate of the variance- ***S*** - may be obtained by the multiplication of the range - ***R*** - by the appropriate factor - ***k*** - that depends on the number of measurements (***N***) of the sample (see table 3) - Eq. 16.:

$$S = k \cdot R \qquad (16)$$

The quick estimate of the P(95%) confidence limit - $\lambda_{95\%}$ of the mean is also possible to derive in the form - Eq.17.:

$$\lambda_{95\%} \cong J \cdot R \qquad (17)$$

where factors - ***J*** - (see table 3.) are also dependent on the degree of freedom of the sample.

Practical remarks

Experimenters very frequently reports the results of measurements in "foggy" way. They simple do not state what they mean by given uncertainty of the result For example : x = 3.25±0.12 ,. the error 0,12 might be: an absolute precision, average deviation, standard deviation, quartile deviation, variance, confidence limits etc. Sometimes it is an estimate based on guesses of what the observer thinks his largest random error might be or an assumption about the possible systematic errors. Frequently it is simply an uneducated guess.
Without careful, clear explanation of the given error values such reports simply have no practical value. At very least one should state clearly what is intended by the error quantity.

Table 3. Factors ***J*** for estimates based on the range ***R***

number ***N*** of data	variance correction factor ***K***	factor ***J*** for 95% confidence
2	0.89	6.40
3	0.59	1.30
4	0.49	0.72
5	0.43	0.51
6	0.40	0.40
7	0.37	0.33
8	0.35	0.29
9	0.34	0.26
10	0.33	0.23

SIGNIFICANT FIGURES AND ROUNDOFF

The precision of the experimental result is implied by the number of digits recorded in the result although the uncertainty should be quoted specifically as well.

The **number of significant figures** in a result is defined as follows.:

1. The most significant digit is the leftmost nonzero digit (e.g.:**2**030,50).
2. The least significant digit is the rightmost nonzero digit if there is no decimal point (e.g.: 2023,**5**).
3. The least significant digit is the rightmost digit (even if it is zero) if there is no decimal point (e.g.: 203**0**).
4. All digits between the least and the most significant digits are counted as significant digits (e.g.: in number 2030 there are 4 sign. digits).

When quoting an experimental result, the number of significant figures should correspond with the uncertainty of the result.

Here are the general rules of the giving significant digits in the result.

The position of the least significant digit is defined by the uncertainty. We may quote the least significant digit only when the uncertainty of the result is smaller (or equal) than 1/2 of this position unit. For example: the result 3.16 means that error is smaller than 0.005 while the result 3.1 means that error is smaller than 0.05.

See table 4. for examples of different number of significant digits in apparently the same result and corresponding uncertainty of it.

Table 4. Significant figures versus error value

Notation	Error smaller or equal to:
24000.00	0.005
24000.0	0.05
24000	0.5
2400 x10	5
240 $x10^2$	50
24 $x10^3$	500
2 $x10^4$	5000

In many cases (e.g. direct experimental data) the number of significant figures may be approximately 1 more than that dictated by the experimental precision. The reason for including this extra digit is to avoid errors that might be caused by rounding errors in later calculations.

For example if the error of the result established in terms of absolute experimental error is about 0.02, while the result of x = 23.54, this result could be quoted as 23.54±0.02 (however when the error is about 0.07 we should probably quote : 23.5±0.07).

Rounding the numbers

When insignificant digits are dropped out from the number the last digit retained should be rounded off for the best accuracy.

Here are the rules of the **arithmetic rounding** off :

1. If the leftmost digit to be dropped is greater than 5 or it is 5 followed by other digits - increase the last retained digit by one.
2. If the leftmost digit to be dropped is less than 5 - leave the last retained digit unchanged.
3. If the leftmost digit to be dropped is exactly 5 - increase the last retained digit by one if it is odd and leave it unchanged if it is even.

For example:

2.34742	-	2.347
2.34751	-	2.348
2.34750	-	2.348
2.34650	-	2.346

Note that with computer analysis of data it is generally advisable to retain the available digits in intermediate calculations and round only the final result.
If sufficient data are available to permit standard deviations and confidence limit to be calculated , there is no particular problem. If the data necessary for these are lacking rules for computations using significant figures help to give a result which express to some degree the uncertainty in the experiment.

When approximate numbers are involved in further calculations the number of significant digits in determined result depends on the number of such digits in the components. It depends on the type of the operation as well :

1. In addition or subtraction - retain only the number of significant digits in the result as it is in the component with the fewest number.

   EXAMPLE :

   32.658 + 11.23 + 4.7 = 48.588 = 48.6

2. In multiplication and division the result should have the relative uncertainty that is the same general magnitude as that in the least precise component. The uncertainty of the result should be in the range of 0.2 to 2 times that of the least precise component.

   EXAMPLE:

   9.32 111 0.038 = 39.31176 = 39

   the least precise component (with the smallest number of significant digits is 0.038. The relative uncertainty is 1/38 = 2.6%.

the relative uncertainty in 39 is 1/39 = 2.5% (while for example in 39.3 it is 0.1/39.3 = 0.25% < 0.2 2.6%)

The experimenter should always keep in mind that any calculation can't improve the certainty of explored results.!

Calculations may be performed by any method that does not introduce round-off errors that are significant in comparison with the experimental errors. The calculations should be organised so as to prevent loss of precision due to the small differences between large quantities. The best way to check a calculation is to repeat it, preferably in a different way, to the same precision.

PROPAGATION OF ERRORS

Experimenter frequently is faced with a task of determining the uncertainty to which is subject a dependent variable - ***F*** - that is a known or tested function of set of independent variables – ***x , y , z*** - subject to errors - ***Δx , Δy , Δz***

The absolute experimental error of the function (single measurements)

The question is : how the uncertainties of measured values ***x_0 , y_0 , z_0*** affect the resulting uncertainties in the final result ***$F_0(x_0, y_0, z_0)$.*** At this point it is assumed that the absolute experimental errors - ***Δx_0, Δy_0, Δz_0*** in the experimental data have been estimated.

The value ***F_0*** is determined by substituting the experimentally determined values ***x_0 y_0 z_0*** of the quantities ***x , y , z*** into a schematically written formula : ***$F_0 = F(x_0, y_0, z_0)$***

Infinitesimal changes ***dx , dy , dz*** in the experimentally value will produce in ***F*** the infinitesimal change ***dF*** - Eq. 18;:

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial z}dz + \ldots \qquad (18)$$

For the finite changes ***Δx , Δy , Δz*** one can have from the Taylor expansion (with only first power terms retained) the approximate formula - Eq.19.:

$$\Delta F = \frac{\partial F}{\partial x}\Delta x + \frac{\partial F}{\partial y}\Delta y + \frac{\partial F}{\partial z}\Delta z + \ldots \qquad (19)$$

Since the ***Δx , Δy , Δz*** represent the absolute experimental errors these errors produce the error ***ΔF*** in the value of ***F***.

The value of the absolute error of the result : ***ΔF_0 (x_0 , y_0 , z_0)*** is then given by the values of the partial derivatives for ***x_0 , y_0 , z_0*** and corresponding experimental errors : ***Δx_0, Δy_0, Δz_0*** - Eq.20.:

$$\Delta F_0 = \left(\frac{\partial F}{\partial x}\right)_{x_0,y_0,z_0}\Delta x_0 + \left(\frac{\partial F}{\partial y}\right)_{x_0,y_0,z_0}\Delta y_0 + \left(\frac{\partial F}{\partial z}\right)_{x_0,y_0,z_0}\Delta z_0 + \ldots \qquad (20)$$

The relative error of the function is then given by the Eq.21.

$$\varepsilon = \frac{\Delta F_0}{F_0} = \frac{1}{F(x_0,y_0,z_0)}\Delta F_0 \qquad (21)$$

This method is also known as a "**differential method**"

Multiply measurements (propagation of random errors)

In the case of multiply measurements of variable values x_0 , y_0, z_0, with already assigned mean - x_0 , y_0, z_0 , the variances $S(x)$, $S(y)$, $S(z)$ (see Eq.10) and confidence limits $\lambda(x)$, $\lambda(y)$, $\lambda(z)$ (see Eq.14), the corresponding variance of the final result F (in terms of normal distribution) is given by - Eq.22.:

$$S^2(F_0)_{\bar{x}_0,\bar{y}_0,\bar{z}_0} = \left(\frac{\partial F}{\partial x}\right)^2_{\bar{x}_0,\bar{y}_0,\bar{z}_0} S^2(x) + \left(\frac{\partial F}{\partial y}\right)^2_{\bar{x}_0,\bar{y}_0,\bar{z}_0} S^2(y) + \left(\frac{\partial F}{\partial z}\right)^2_{\bar{x}_0,\bar{y}_0,\bar{z}_0} S^2(z) \qquad (22)$$

The final result F_0 is calculated using mean values x_0 , y_0, z_0 for the observed quantities. The confidence limit of final result $\lambda [F(x_0 , y_0, z_0)]$ is given by the confidence limits of each variable $\lambda(x)$, $\lambda(y)$, $\lambda(z)$ in Eq.23.:

$$\lambda^2[F_0(\bar{x}_0,\bar{y}_0,\bar{z}_0)] = \left(\frac{\partial F}{\partial x}\right)^2_{\bar{x}_0,\bar{y}_0,\bar{z}_0} \lambda^2(x) + \left(\frac{\partial F}{\partial y}\right)^2_{\bar{x}_0,\bar{y}_0,\bar{z}_0} \lambda^2(y) + \left(\frac{\partial F}{\partial z}\right)^2_{\bar{x}_0,\bar{y}_0,\bar{z}_0} \lambda^2(z) \qquad (23)$$

Equations (22,23) are exact for linear functions of the variables and hold satisfactorily for other functions if the deviations of x , y , z etc. are less than about 20% of the mean values x_0 , y_0, z_0.
Examples for different types of functions are given below -Eq.24.:

(24)

1.

$$F = ax \pm by \pm cz$$
$$S^2(F) = a^2S^2(x) + b^2S^2(y) + c^2S^2(z)$$
$$\lambda^2(F) = a^2\lambda^2(x) + b^2\lambda^2(y) + c^2\lambda^2(z)$$

2.

$$F = axyz \qquad or:(axy/z) \quad or:(ax/yz) \quad or:(a/xyz)$$
$$S^2(F) = F^2 \cdot \left[\frac{1}{x^2}S^2(x) + \frac{1}{y^2}S^2(y) + \frac{1}{z^2}S^2(z)\right]$$

$$\lambda^2(F) = F^2 \cdot [\frac{1}{x^2}\lambda^2(x) + \frac{1}{y^2}\lambda^2(y) + \frac{1}{z^2}\lambda^2(z)]$$

3.

$$F = ax^b$$

$$S^2(F) = F^2 \cdot [b^2 \cdot \frac{S^2(x)}{x^2}]$$

$$\lambda^2(F) = F^2 \cdot [b^2 \cdot \frac{\lambda^2(x)}{x^2}]$$

4.

$$F = a \cdot \exp(\pm bx)$$

$$S^2(F) = F^2 \cdot [b^2 \cdot S^2(x)]$$

$$\lambda^2(F) = F^2 \cdot [b^2 \cdot \lambda^2(x)]$$

5.

$$F = a^{\pm bx}$$

$$S^2(F) = F^2 \cdot (b \cdot \ln a)^2 \cdot S^2(x)$$

$$\lambda^2(F) = F^2 \cdot (b \cdot \ln a)^2 \cdot \lambda^2(x)$$

6.

$$F = a \cdot \ln(\pm bx)$$

$$S^2(F) = a^2 \frac{S^2(x)}{x}$$

$$\lambda^2(F) = a^2 \frac{\lambda^2(x)}{x}$$

When the expression for ***F*** is complicated one it is advisable to watch for opportunities for transforming the expression or parts of it so, that single symbols representing known quantities can be substituted for groups of several symbols.
The most advantageous simplifications are usually obtained when the quantities substituted are the end results of the calculations such as ***F*** itself, or at least quantities, which represent the later stages of the calculation of ***F***. It gives the most frequent opportunities for cancellation and simplification.

LEAST SQUARES METHOD FOR LINEAR APPROXIMATION

Frequently the experimenter wishes to determine the function ***y = y(x)*** that describes the relation between two measured variables given by the set of ***N*** pair (x_i, y_i $i=1, 2, 3, N$) .
A sufficient number of numerical methods is related with this problem. By means of them we can construct even complex functions to fit the data closely, such a polynomials, exponential, non-linear functions etc.

Particular analytic form of dependence may be suggested either by the characteristics of the phenomena (theoretical predictions) or the assumption based on obtained data distribution (e.g. graphs).
The analytic expressions usually depend on a few parameters and establishing of these parameters is the main task of numerical method involved. Having the analytic expression we may subsequently differentiate, integrate it etc. by analytic methods without further need for numerical work.

The least-square fit to a straight line is the most frequent procedure used in students laboratories.

Let's consider the problem of pairs of variables (x_i , y_i) that are functionally related to one another. So, we suppose that the experimental values y_i are expected to conform to analytic functional form - Eq. 25.:

$$y = f(x, a_0, a_1, a_2, ...) \qquad (25)$$

The most probable values for the coefficients a_k are to be determined.
To draw a line through all data points is obviously impossible but we can try to find value of the parameters a_k that minimise the discrepancy between the measured values y_i and calculated values $y(x_i)$. We cannot determine those parameters exactly but try to find the most probable estimates for the coefficients.

The most frequently used method for this purpose is the least square method based on the normal error distribution .

The criteria for the minimising the discrepancy between the measured and predicted values y_i in this method is to minimise the sum of the squares of the deviations - S - Eq. 26.

$$S = \sum_{i=1}^{N} \left[y_i - f(x_i, a_0, a_1, a_2, .. a_k \right]^2 \qquad (26)$$

The number - k - of parameters a_i should be considerably smaller than the number N of the experimental values. It is not required that the x_i be equally spaced.

Since we are looking for the polynomial - Eq.27.:

$$f(x_i, a_0, a_1, ... a_k) = \phi_0(x)a_0 + \phi_1(x)a_1 + ... \phi_k(x)a_k \qquad (27)$$

the Eq.26. could be written in the form - Eq.28:

$$S = \sum_{i=1}^{N} \left[y_i - \phi_0(x_i)a_0 - \phi_1(x_i)a_1 - ..... \phi_k(x_i)a_k \right]^2 \qquad (28)$$

The minimisation of - S requires that - Eq.29.:

$$\frac{\partial S}{\partial a_0} = \frac{\partial S}{\partial a_1} = \ldots\ldots = \frac{\partial S}{\partial a_k} = 0 \qquad (29)$$

A set of equations -Eq.29 - after the expanding ***f*** in the Taylor series and keeping only terms to first order it become to a set of the equivalent equations called the normal equations .
For the linear approximation i.e. for a function : ***y = a+bx*** the corresponding expressions for parameters - ***a*** and ***b*** are given by the Eq. 30.:

$$a = \frac{1}{D}\cdot\begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i\, y_i & \sum x_i^{\,2} \end{vmatrix} = \frac{1}{D}\left(\sum_{i=1}^{N} x_i^{\,2}\cdot\sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i\, y_i\right)$$

$$b = \frac{1}{D}\begin{vmatrix} N & \sum y_i \\ \sum x_i & \sum x_i\, y_i \end{vmatrix} = \frac{1}{D}\left(N\cdot\sum_{i=1}^{N} x_i\, y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i\right)$$

$$D = \begin{vmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^{\,2} \end{vmatrix} = N\cdot\sum_{i=1}^{N} x_i^{\,2} - \left(\sum_{i=1}^{N} x_i\right)^2$$

(30)

This result is obtained for a special case in which all the uncertainties (standard deviations) of the measured variables - y_i i.e. σ_i are equal.
If not, the general solutions for parameters ***a , b*** are - Eq.31:

$$a = \frac{1}{D}\cdot\begin{vmatrix} \sum \frac{y_i}{\sigma_i^{\,2}} & \sum \frac{x_i}{\sigma_i^{\,2}} \\ \sum \frac{x_i\, y_i}{\sigma_i^{\,2}} & \sum \frac{x_i^{\,2}}{\sigma_i^{\,2}} \end{vmatrix}$$

$$b = \frac{1}{D}\cdot\begin{vmatrix} \sum \frac{1}{\sigma_i^{\,2}} & \sum \frac{y_i}{\sigma_i^{\,2}} \\ \sum \frac{x_i}{\sigma_i^{\,2}} & \sum \frac{x_i\, y_i}{\sigma_i^{\,2}} \end{vmatrix}$$

$$D = \begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{vmatrix} \qquad (31)$$

Error of the least -square- fit

When the uncertainties of measurement are instrumental and all the data are recorded with the same precision (as it frequently happens), we can obtain a common error in measurements of ***y*** from the obtained fit. The estimate standard deviation of as individual measurement $\sigma = \sigma_I$ is given by the Eq.32.:

$$\sigma^2 \cong \frac{1}{N-2} \cdot \sum_{i=1}^{N} \left(y_i - a - b\, x_i \right)^2 \qquad (32)$$

In order to find the uncertainty in the estimation of the established parameters ***a , b*** of the fit, the propagation of the errors should be discussed.

For the special case of common uncertainties in y_i ($\sigma_I = \sigma_I$) standard deviations of the obtained parameters ***a , b*** are given by the Eq.33.:

$$\sigma_a^2 = \frac{\sigma^2}{D} \cdot \sum_{i=1}^{N} x_i^2$$

$$\sigma_b^2 = N \cdot \frac{\sigma^2}{D} \qquad (33)$$

where:

σ - is given by the Eq.32.

D - is given by the Eq.30.

The strength of relation between two variables ***x, y*** can be estimated by frequently used in calculus of probability so called " correlation coefficient" - ρ - defined by the Eq.34.:

$$\rho = \frac{1}{N-1} \cdot \sum \frac{\left(x_i - \bar{x} \right) \cdot \left(y_i - \bar{y} \right)}{S_x \cdot S_y} \qquad (34)$$

where :

S_x , S_y are variances of variables ***x, y*** given by the Eq.10.

The correlation coefficient ρ values are from the range $-1 < \rho < 1$. For $\rho = 1$, there is no relation between variables while for $\rho = \pm 1$ there is an ideal correlation between ***x*** and ***y*** . Positive values of correlation coefficient are related to the so called "positive correlation", i.e. while increasing one of the variables the other one

subsequently increases too. Negative correlation coefficient is related to the relation in which increase of one variable leads to the decrease of the other.

In the Fig. 2. there are shown several examples of data distribution, the linear fit and corresponding values of the correlation coefficient.

The correlation coefficient itself does not yet give the clear information about the strength of the relation. It always should be related to the number of measurements (or rather degree of freedom - for linear approximation = N-2). The critical (minimal) values of correlation coefficient versus the number of measurements ***N*** are given in the table 5, for different values of ***P%*** confidence.

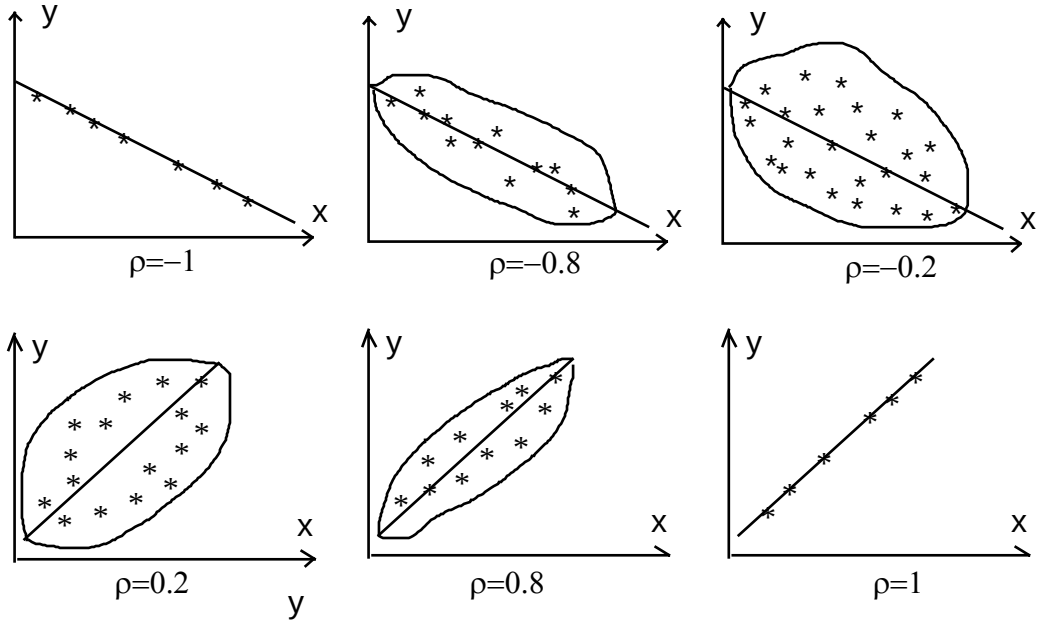

Fig.2 . Linear - fit for different correlation coefficient.

Table 5. Critical values of correlation coefficient

number of data	critical correlation coeff. for $P_{\%}$ confidence			
	90%	95%	98%	99%
1	0.98769	0.99692	0.999507	0.999877
2	0.90000	0.95000	0.980000	0.990000
3	0.8054	0.8783	0.93433	0.95873
4	0.7293	0.8114	0.8822	0.91720
5	0.6694	0.7545	0.9329	0.8745
6	0.6215	0.7068	0.7887	0.8343
7	0.5822	0.6664	0.7492	0.7977
8	0.5494	0.6319	0.7155	0.7646
9	0.5214	0.6021	0.6851	0.7348
10	0.4973	0.5760	0.6581	0.7079
11	0.4762	0.5529	0.6339	0.6835
12	0.4575	0.5324	0.6120	0.6614
13	0.4409	0.5139	0.5923	0.6411

14	0.4259	0.4973	0.5742	0.6226
15	0.4124	0.4821	0.0.5577	0.6055
16	0.4000	0.4683	0.5425	0.5897
17	0.3887	0.4555	0.5285	0.5751
18	0.3783	0.4438	0.5155	0.5614
19	0.3687	0.43290	0.5034	0.5487
20	0.3598	0.4227	0.4921	0.5368

APPENDIX.

Tables.

Table I

Physical quantities and their units

Quantity	Unit (SI)	Symbol
length	meter	m
time	second	s
frequency	Hertz	Hz
mass	kilogram	kg
density	kilogram per cubic meter	kg/m^3
work, energy	Joule	J
	electron-volt	eV
power	Watt	W
electric charge	coulomb	C
electric current	Ampere	A
electric field	Volt per meter	V/m
potential difference voltage	Volt	V
electric resistance	Ohm	Ω
resistivity	Ohm meter	Ω m
capacitance	Farad	F
magnetic flux	Weber	Wb
magnetic field	Tesla	T
inductance	Henry	H
luminous intensity	candela	cd
radiant intensity	Watt per steradian	W/sr
illumination	Lux	lx
flux	Lumen	lm

Table I I

Physical constants

Constant	Symbol	Value
elementary charge	E	$1.602\ 10^{-19}$ C
electron rest mass	M_e	$9.11\ 10^{-31}$ kg
Planck constant	H	$6.63\ 10^{-34}$ J s
Stefan-Boltzmann constant	σ	$5.67\ 10^{-8}$ W/m^2 K^4
Boltzmann constant	K	$1.38\ 10^{-23}$ J/K
permittivity constant	ε_0	$8.85\ 10^{-12}$ F/m
permeability constant	μ_0	$1.26\ 10^{-6}$ H/m

Table I I I

Spectrum of visible light versus colours

wavelength [nm]	corresponding colour
400 - 450	Violet
450-500	Blue
500-550	Green
550-590	Yellow
590-640	Orange
640-700	Red

Table I V.

Spectral lines of gasses and metal vapours

System	wavelength of line spectrum [nm]
hydrogen (H)	656.2 ; 486.1 ; 434.0 ; 410.2 ;
helium (He)	706.5 ; 667.8 ; 587.6 ; 501.6 ; 486.6;
neon (Ne)	703.2; 629.9; 588.2; 585.3; 540.1; 534.1; 497.7; 488.5; 482.7; 475.3; 470.4;
sodium (Na)	617.2; 588.4; 567.7; 514.1; 498.7;
mercury (Hg)	671.6; 579.0; 577.0; 546.0; 435.8;
He-Ne laser	632.8

Table V

SJ Units Prefixes

Prefix	Symbol	Factor
exa	E	10^{18}
peta	P	10^{15}
tera	T	10^{12}
giga	G	10^{9}
mega	K	10^{6}
kilo	K	10^{3}
hecto	H	10^{2}
deca	Da	10^{1}
deci	D	10^{-1}
centi	C	10^{-2}
mili	M	10^{-3}
micro	□	10^{-6}
nano	N	10^{-9}
pico	P	10^{-12}
femto	F	10^{-15}
atto	A	10^{-18}