Appendix to the Physical-chemical practical course

1st semester

Part I

Guidelines

for the preparation of scientific protocols
in the physical-chemical practical course

Part II

Introduction to error calculation

Part I: Recording and evaluation of measurements

1) Measurement protocol:

During the experiment, each group prepares a measurement protocol, which should be understandable, legible and not written in pencil.

It should contain:

- Test number and name of the test
- Group number and names of the persons carrying out the experiment
- Date the experiment was carried out
- Measurement results in tabular form with units of the measured variables

The measurement protocol must be signed off by an assistant (supervisor of the experiment, or by another person if the supervisor is absent) after completion of the experiment!!!

The measurement protocol is a strict component of the test protocol to be prepared!

2) Test protocol:

For each experiment, a protocol is prepared by the group, which is to be filed in the storage boxes provided for this purpose no later than one week after the experiment has been carried out and returned to the same place after being reviewed by the assistant. The report is then either marked “tested” or “consultation”.

Consultation means:

- the protocol will not be accepted in its current form
- the respective assistant must be contacted as soon as possible (group)
- the minutes must be redrafted, errors corrected and resubmitted
- the report is normally given a poor grade (5)
- the process is repeated until the minutes are certified
- outstanding consultations and ultimately untested minutes are assessed as work not done.

(Do not correct without prior consultation! The assistant will tell you what to do)

The entire group is responsible for the protocol!

The experimental protocol should be structured as follows:

Number and name of the trial

Name:

Group No:

Date:

I. Aim of the experiment:

What is to be shown (1-2 sentences)

II. Theory of the experiment:

Physico-chemical measurements are always based on physical laws and theories.

The theory on which the experiment is based and the laws used to evaluate the data are presented and explained here.

This should be done in a short, concise form and should focus on the most important (necessary for understanding the data analysis).
Historical treatises are just as out of place here as transcripts of experimental instructions or textbook pages.

III Experiment setup and execution:

The procedure of a measurement is explained by means of a sketch (no freehand drawings) of the experimental setup.

Avoid monotonous descriptions of pure action sequences such as “We opened tap #1 and then read off the value. Then we wrote down the corresponding value and started with the 2nd measurement by ...” should be avoided at all costs !!!!!

Important: A protocol is not an instruction manual for carrying out the experiment!

Never write protocols in the “I” or “we” form!
The measured values are presented in a table !

IV. Evaluation:

In accordance with the laws set out in II, the graphical and numerical evaluation of the measured values is carried out here (tables in III).

All the relationships and equations used are given here.

All results, including those of intermediate calculations, are summarized in tables wherever possible. A detailed, complete calculation of the result is presented and carried out using the example of a measurement. The values are given with a reasonable number of digits resulting from the measurement accuracy (see error consideration), the error and its dimensions.

Graphical representations are very important for physical measurements and should be carried out whenever possible (see experiment instructions).

Preparation of a graphical representation:

The pairs of values, determined from direct measurement or calculated results, are entered on graph paper in the rectangular coordinate system and connected by a smooth curve with a fine pencil.

Please note that:
- curved curves are drawn out with a curve ruler, individual sections merge continuously into one another
- curves are drawn through scattering measured values so that these are distributed equally on both sides
- If individual points deviate strongly from the course of the curve, this is a signal for measurement errors and possible outliers (discussion!)
- the scale should be chosen so that the curve is not distorted and the dependence of the value pairs remains recognizable (not too steep, not too flat)
- a unit length (e.g. 1 °C) must be represented by an even number (e.g. 10, 20, 50 mm) (not 11.2 mm)
- it is not necessary to choose the same scale for both axes and to start the coordinate system at the origin
- the measured values must be provided with error bars
- the axes are labeled and provided with corresponding units
(e.g. pressure p in [Pa] , or temperature T in [K])

V. Discussion of the results and error analysis:

To what extent does the result correspond to the expected value?
(Comparison with literature)

How can the result be assessed qualitatively and quantitatively?

Qualitatively, a test result can be assessed by asking the following questions:
- How accurate is the reading and how does this influence the result?
- What possible sources of error are present and what effects do they have ?
- Which measurements must be carried out with particular care, which not and why?
- Which external conditions (e.g. temperature) affect the result and how?

Discuss possible deviations of individual measured values !
Discuss possible deviations of your values from literature values !

A quantitative assessment of your result is made by means of an error calculation!

A measurement result without indication of its error limits is not a result!

VI. original measurement protocol with the assistant's certificate !!!!!

Part II: Introduction to error calculation

All results of physico-chemical measurements are subject to errors and must therefore be stated, for example, in the form T = 112.3 ± 4.7 K

The error to be specified results either from an error calculation or a graphical application.

Errors are deviations of the measured value from an unknown error-free result. This “true measured value” can be determined approximately by calculating the arithmetic mean ($\bar{x}$) from n individual measurements (x_i). The accuracy increases with the number of individual measurements.

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

Errors are divided into 2 groups:

Systematic errors:

They change the result in one direction (larger or smaller). They are caused by deficiencies in the measuring method (incorrect calibration, impure chemicals) or by the way the experimenter works, e.g. systematically incorrect reading.

I.e: For each experiment, consider whether systematic errors occur and how they can be eliminated at the beginning of the experiment.

Random errors:

Fluctuations in individual results in both directions (sometimes too large, sometimes too small) are unavoidable! According to Gauss's law of error distribution (see above), the fluctuations are distributed in such a way that the values around the mean value occur most frequently and the number of measured values decreases at a greater distance from it (bell curve).

Error of individual measurements:

Absolute error: Difference between measured value x_i and mean value $\bar{x}$ $dx = x_i - \bar{x}$

Relative error: Ratio of the absolute error to the mean value $dx / \bar{x}$
Percentage error: relative error * 100, the accuracy of the measured value x_i is $dx/\bar{x}$ *100 %.

Statistical error effects:

The mean square error σ_X of the individual measurements is a measure of the dispersion of the measured values around the mean value, i.e. a measure of the quality of the experimental method:

Standard deviation of the individual measurements: $\sigma_x = \pm\sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}$

The mean square error $\sigma_{\bar{X}}$ of the mean value is a measure of the accuracy of the mean value (the result)

Standard deviation of the mean value: $\sigma_{\bar{x}} = \pm\sqrt{\frac{\sigma_x{}^2}{n}} = \pm\sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n*(n-1)}}$

Example: Measurement of viscosities:

$t_1 = 231.7$ s $|\ t_1 - \bar{t}\ |^2 = |231.7 - 231.929|^2\ s^2 = 0.052\ s^2$

$t_2 = 231.9$ s $|\ t_2 - \bar{t}\ |^2 = |231.9 - 231.929|^2\ s^2 = 0.001\ s^2$

$t_3 = 232.2$ s $|\ t_3 - \bar{t}\ |^2 = |\ 232.2 - 231.929|^2\ s^2 =$

$t_4 = 231.6$ s ....

$t_5 = 231.8$ s ....

$t_6 = 232.0$ s ....

$t_7 = 232.3$ s ....

$\bar{t} = 231.929$ s $\sum_{i=1}^{n} \left|\ t_i\ -\ \bar{t}\ \right|^2 = 0.395\ s^2$

Standard deviation of the individual measurements: $\sigma_t = \pm\sqrt{\frac{0.395}{7-1}} = \pm 0.256$ s

mean error of the mean value: $\sigma_{\bar{t}} = \pm\sqrt{\frac{0.395}{(7-1)*7}} = \pm 0.097$ s

Final result: $\bar{t} = 231.9 \pm 0.1$ s (1σ)

Error propagation:

In the previous section, methods for specifying errors of individual measurements were discussed. However, a measurement result is often obtained from several individual measurements, each of which can be subject to different errors.

Example: To determine a speed v, a distance s is measured, as well as the time t required for this. The equation is then: v = s / t.

The distance measurement is subject to an error Δs, the time measurement is subject to an error Δt. The error of the speed v is therefore dependent on both measurement errors.

The maximum speed is: largest possible numerator / smallest possible denominator, i.e. $v_{max} = s + \Delta s / t - \Delta t$. (Minimum speed: $v_{min} = s - \Delta s / t + \Delta t$)

For more complicated formulas and generally considered and mathematically correct, the maximum error results from an error propagation of the individual errors.

Given a function z = f(x,y,...), the maximum error of z is given by :

$$\Delta z = \left|\frac{\partial z}{\partial x} * \Delta x\right| + \left|\frac{\partial z}{\partial y} * \Delta y\right| + \cdots$$

Where Δx and Δy are the errors of the individual measurements of x and y, $\frac{\delta z}{\delta x}$ and $\frac{\delta z}{\delta y}$ the partial derivatives of the function z with respect to x and y.

However, the maximum error is usually not made. It is therefore important to specify an average error, which is calculated as follows:

$$\Delta z = \sqrt{\left(\frac{\delta z}{\delta x}\right)^2 * \Delta x^2 + \left(\frac{\delta z}{\delta y}\right)^2 * \Delta y^2 + \ldots.}$$

Example: Determination of a volume v_3 via the ideal gas law using 3 pressure measurements and a known initial volume v_1,

$$v_3 = p_1 * v_1 * \left(\frac{1}{p_3} - \frac{1}{p_2}\right)$$

whereby the measured values p_1, p_2 and p_3 are each subject to a reading error of ± 1 Torr.

$$\Delta v_3 = \sqrt{\left(\frac{\delta v_3}{\delta p_1}\right)^2 * \Delta p_1^2 + \left(\frac{\delta v_3}{\delta p_2}\right)^2 * \Delta p_2^2 + \left(\frac{\delta v_3}{\delta p_3}\right)^2 * \Delta p_3^2}$$

$$\Delta v_3 = \sqrt{\left(\frac{v_1}{p_3} - \frac{v_1}{p_2}\right)^2 * 1^2 + \left(\frac{p_1 * v_1}{p_2^2}\right)^2 * 1^2 + \left(-\frac{p_1 * v_1}{p_3^2}\right)^2 * 1^2}$$

Compensation line, method of least squares:

A measured variable is often measured as a function of a second variable. In many cases this relationship is linear,

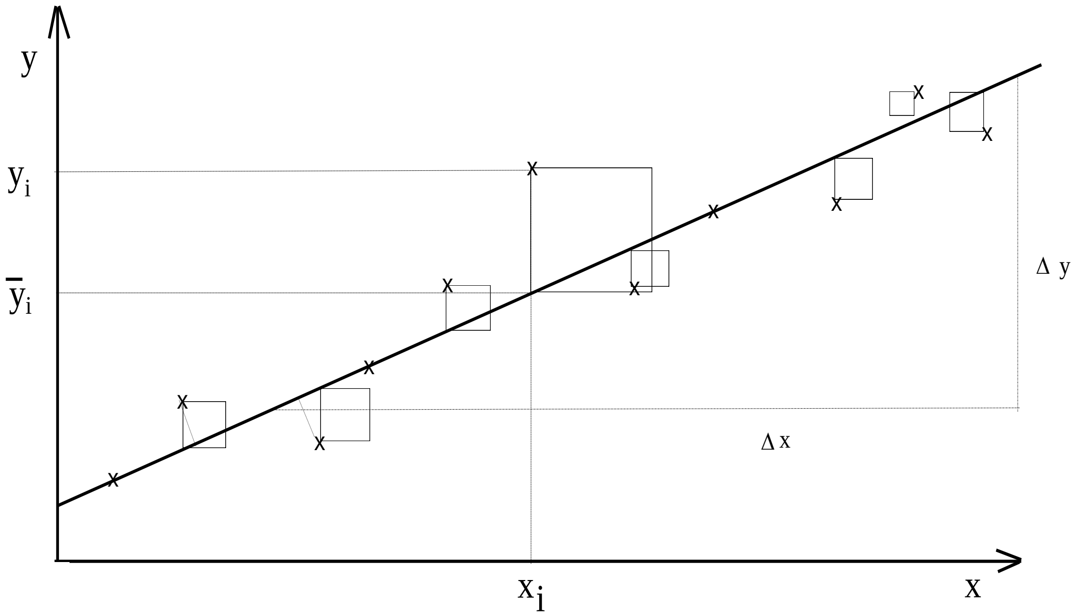
Linear equation: y = m * x + b Slope m = $\frac{\Delta y}{\Delta x}$

or can be linearized: $y = a * x^n \Rightarrow \lg y = \lg a + n * \lg x$

if lg y is now plotted against lg x, the slope of the straight line gives a value for n, the value a follows from the intercept lg a.

The problem lies now in determining the “most ideal” straight line, the one that best describes the sum of the measured values xi, yi.

Mathematically, the problem is solved by minimizing the error sum of squares:



Measured values = x_i, y_i ; “true” values = x_i, $\overline{y_i}$; error = $y_i - \overline{y_i}$

Requirement: Function f (m,b) = $\sum_i (y_i - \overline{y_i})^2 = \sum_i (y_i - m*x_i - b)^2 \rightarrow \text{Minimum}$

i.e.: f' (m) = 0; f' (b) = 0

The 1st derivative of the function with respect to m and b is formed and set equal to 0, resulting in 2 equations with 2 unknowns which can be solved for m and b:

$$m = \frac{n*\sum_i x_i y_i - \sum_i x_i * \sum_i y_i}{n*\sum_i x_i^2 - (\sum_i x_i)^2} \qquad b = \frac{\sum_i y_i * \sum_i x_i^2 - \sum_i x_i * \sum_i x_i y_i}{n*\sum_i x_i^2 - (\sum_i x_i)^2}$$

If the sums Σx_i, Σy_i, $\Sigma x_i y_i$, und Σx_i^2 are formed from the measured values, m and b can be calculated.

The errors of m and b, Δm and Δb can also be easily calculated:

Error sum of squares S : $S = \sum_i (m*x_i + b - y_i)^2$

standard deviation: $\sigma_y = \sqrt{\frac{S}{n-2}}$

$$\Delta m = \sigma_y * \sqrt{\frac{n}{n*\sum_i x_i^2 - (\sum_i x_i)^2}} \qquad \Delta b = \sigma_y * \sqrt{\frac{\sum_i x_i^2}{n*\sum_i x_i^2 - (\sum_i x_i)^2}}$$

so that the values can be given as m ± Δm, or b ± Δb.

Another important parameter is the correlation coefficient r. It indicates how well the measuring points correspond to a straight line:

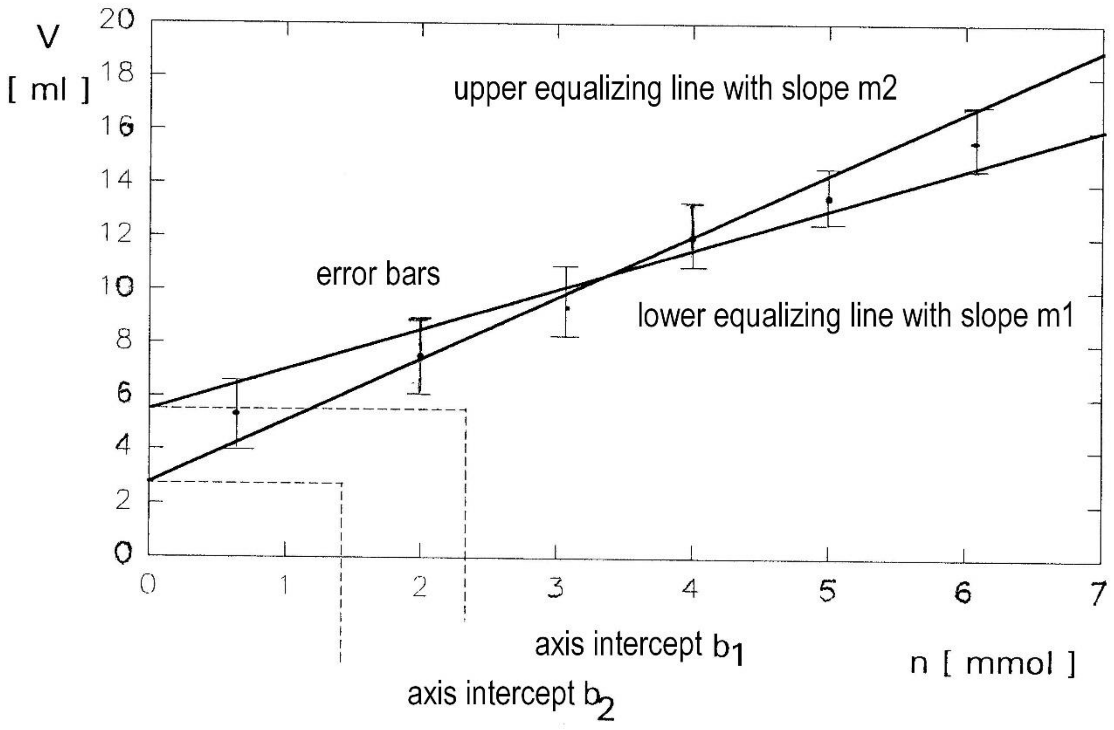
$$r = \frac{(n * \sum_i x_i y_i - \sum_i x_i * \sum_i y_i)^2}{(n * \sum_i x_i{}^2 - (\sum_i x_i)^2) * (n * \sum_i y_i{}^2 - (\sum_i y_i)^2)}$$

r = 0 : no correlation

r = 1 : complete correlation $0 \leq r \leq 1$!!!!

Error data from graphical evaluation:

In a graphical plot, the estimated experimental uncertainties of the values are shown as error bars of the values in the diagram!
Two straight lines are drawn through the error bars and the corresponding slopes (minimum and maximum) and the corresponding axis intercepts are determined.



The “true” values m and b are obtained by arithmetic averaging.
The errors Δm and Δb are obtained by halving the differences.

m = 0.5 * (m2 + m1) ± 0.5 * (m2 - m1) [ml / mmol]

b = 0.5 * (b_2 + b_1) ± 0.5 * (b_2 - b_1) [ml / mmol]

Gaussian error distribution function

As already mentioned before, the fluctuations of the measured values around the mean value, i.e. the frequency φ (v) of random errors of the magnitude v ($x_i - \bar{x}$) are subject to a distribution law. The function must satisfy the following requirements:

1.) $\varphi \geq 0$; $\varphi (+v) = \varphi (-v)$

The frequencies φ are always positive, i.e. the curve lies above the x-axis, positive and negative deviations are equally probable, i.e. the curve is symmetrical to the y-axis

2) $\varphi (v1) > \varphi (v2)$ if $| v_1 | < | v_2 |$;

small deviations are more common than large errors

3.) $\varphi_{max} = \varphi (0)$ and $\varphi_{min} = \varphi (\pm\infty)$

The function has a maximum at an error of 0, infinitely large errors do not occur.

4) φ must quickly become small for large errors, i.e. the curve quickly approaches the x-axis

A function that fulfills all these requirements is a “bell curve”:

$y = k * e^{-a * x^2}$ or $\varphi(v) = C_1 * e^{-C_2 * v^2}$ with $C_1, C_2 > 0$

The shape of this curve is determined by the constants C1 and C2. The height of the maximum is determined by the number and accuracy of the measured values, i.e. the standard deviation is included in the constants C_1 and C_2.

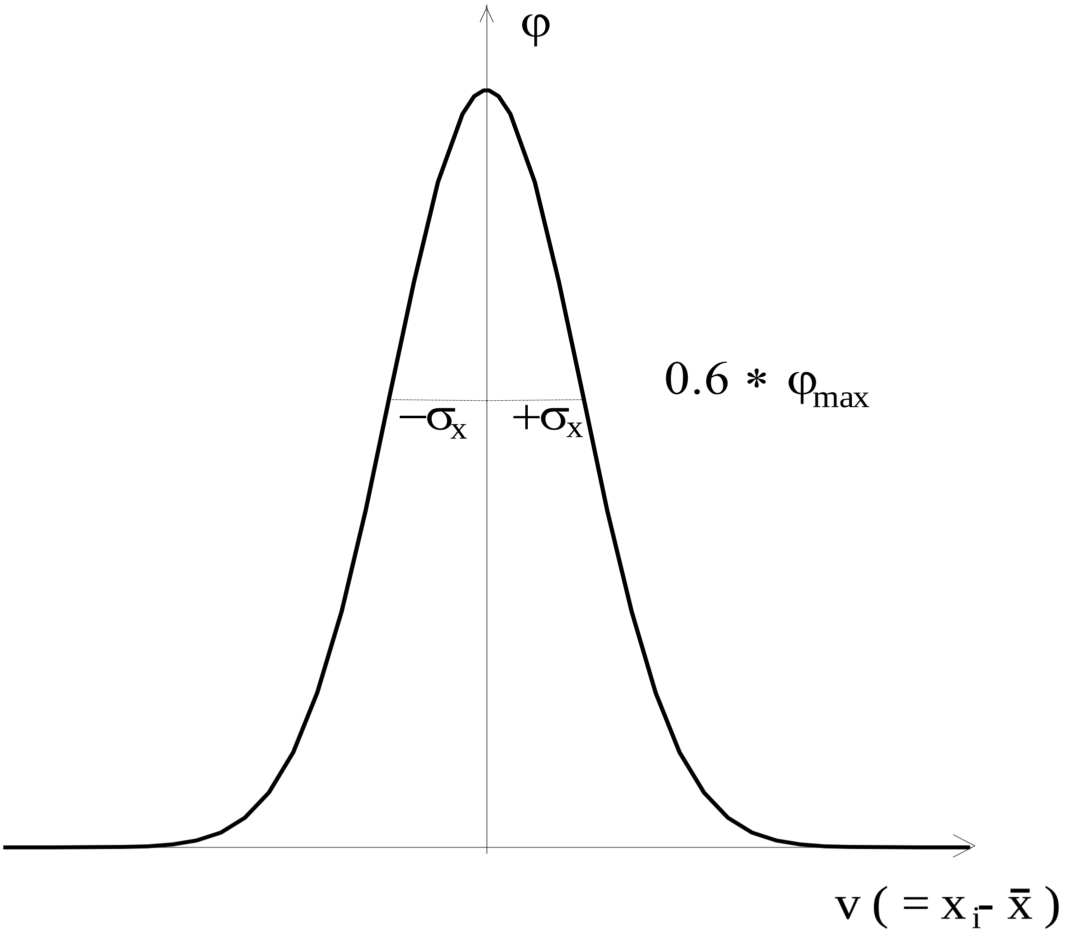
$$\sigma_x = \pm\sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} \quad , C_1 = \frac{h}{\sqrt{\pi}} = \frac{1}{2\pi * \sigma_x} \quad \text{and} \quad C_2 = h^2 = \frac{1}{2 * \sigma_x}$$

The value h is also referred to as the dispersion number or accuracy number.
For v=0 it follows (since $e^0 = 1$) that the frequency φ (0) is equal to C_1 (maximum).

The two inflection points of the curve are at: $\varphi(v_w) = \frac{1}{h * \sqrt{2}} = \pm\sigma_x$

The heights of the inflection points relative to the maximum are given by:

$$\frac{\varphi(v = \sigma_x)}{\varphi(v = 0)} = \frac{\frac{1}{\sqrt{2\pi} * \sigma_x} * e^{-\frac{1}{2\sigma_x} * {\sigma_x}^2}}{\frac{1}{\sqrt{2\pi} * \sigma_x} * e^{-\frac{1}{2\sigma_x} * 0^2}} = e^{-\frac{1}{2}} = 0.6056$$



The probability that a measured value lies within a certain error interval Δv can be calculated by integrating the curve.

The total area under the curve is 100 %; the area for $\bar{x} \pm 1\sigma_x = 68.3$ %
$\bar{x} \pm 2\sigma_x = 95.4$ %
$\bar{x} \pm 3\sigma_x = 99.7$ %

The standard deviation is therefore a quality mark for the measurement method used.
An indication $\bar{x} = 112.30 \pm 1.25$ ($2\sigma_x$) means that a measured value x_i lies with a 95% probability within a scatter range of $\pm 2\sigma_x$ around the "true" value.

A value is an **"outlier"** if $x_A - \bar{x} \geq k * \sigma_x$ (k=3 or 4). (Determination of $\bar{x}$, σ_x without x_A).

Other distribution functions:

The Gaussian error distribution function is a statistical method and, strictly speaking, only applies to a very large number of measured values. However, since in reality only a limited number of measured values are determined, other distribution functions, so-called sampling distributions, are required in order to be able to indicate with a certain statistical certainty S whether a determined value lies within a certain confidence limit. The most frequently used distribution is the Student- or t-distribution, which is shown in the following table.

f	S = 95%	S = 99%
1	12.71	63.66
2	4.30	9.92
3	3.18	5.84
4	2.78	4.60
5	2.57	4.03
6	2.45	3.71
7	2.37	3.50
8	2.31	3.36
9	2.26	3.25
10	2.23	3.17
12	2.18	3.06
15	2.13	2.95
18	2.10	2.88
20	2.09	2.85
25	2.06	2.79
30	2.04	2.75
35	2.03	2.72
40	2.02	2.70
50	2.01	2.68
100	1.984	2.626
200	1.972	2.601
300	1.968	2.592
∞	1.960	2.576

f = n-1 = number of degrees of freedom

The values are to be interpreted as follows:

With the statistical certainty S, the mean value $\overline{X}$ lies within the confidence limit of $\overline{X} \pm t_S * \sigma_X$.

For $n \to \infty$ the t-distribution function changes to the Gaussian distribution function.

For our example of viscosity measurements the following applies:

n = 7, f = 6, i.e. t_s (99%) = 3.71 (see table)

The measurement resulted in:

$\overline{X} = 231.93 \quad \sigma_X = \pm 0.256 \quad \sigma_{\overline{X}} = \pm 0.097$

With a probability of 99%, the “true” value is therefore within an error interval of:

$\overline{X} = 231.93 \pm 0.36$ [s]. (from 3.71*0.097)

With a probability of 99%, any measured value x_i of the measurement series lies within the interval of

$\overline{X} = 231.93 \pm 0.95$ [s]. (3.71*0.256)

This script is intended to serve as an introduction to error calculation by explaining the basic principles of estimating measurement errors in physico-chemical measurements in a simple way. It makes no claim to completeness; on the contrary, the explanation of a large number of undoubtedly important things such as scatter, variance, weightings, sensitivities, etc. has been deliberately omitted.

“The errors in this script are subject to a normal distribution. (2σ)”.

K.D. Setzer