ERROR ANALYSIS

NOTATION

When reporting results it is not sufficient to just give the value found, also the error or uncertainty has to be reported. This is usually written as:

$$x \pm \delta x$$

Where x is the result of the measurement and δx is the uncertainty. Generally, δx is given with one significant digit and the result of the measurement is given with a similar accuracy.

For example:	$A = 93.14 \pm 0.03$	$B = (12.67 \pm 0.06)\ 10^3$
	$A = 93.1 \pm 0.5$	$B = (12.6 \pm 0.2)\ 10^3$
	$A = 93 \pm 2$	$B = (12 \pm 2)\ 10^3$
thus **NOT**	~~$A = 93.179 \pm 0.59874$~~	

However, when the uncertainty starts with a 1 then usually two significant digits are given.

For example:	$A = 93.14 \pm 0.13$	$B = (12.67 \pm 0.16)\ 10^3$
	$A = 93.1 \pm 1.5$	$B = (12.6 \pm 1.2)\ 10^3$

It speaks for itself that the results are given with the appropriate SI units.

For example, if we have measured the mass of a piece of metal we write:

$$m = 12.2 \pm 0.3 \text{ kg}$$

DETERMINATION VALUE AND UNCERTAINTY

It is not always possible to obtain an idea about the accuracy of a quantity by just one measurement. If we for example want to measure the vibrational period of a mass attached to a string we can do this using a stopwatch. The accuracy with which we can measure is determined by the resolution of the watch and of course the accuracy by which we start and stop the watch. The first quantity is easily determined, the second is less well defined. We either have to estimate this value, which is not always easy, or repeat the experiment several times to get an idea about the accuracy. If we perform the experiment several times we will find a distribution of results. From this distribution we can derive the best value for the measurement and the appropriate uncertainty using statistical theories.

If we perform the measurement N times and find the results (x_1, x_2, x_3,… x_N) than the value that describes the real value the best is given by the average of measured results:

$$x_{best} = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

We define the standard deviation, σ, as:

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

The standard deviation is the interval around the average value within which 68.3% of all the measurements fall. Using the standard deviation we can deduce the accuracy with which the average value is known (also known as the standard deviation of the average) by making use of the following relation:

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

Whereas the standard deviation does not change with the number of measurements the standard deviation of the average gets better with increasing number of measurements.

Lets have a look at the following example. A student measures the vibrational period of a mass suspended from a spring using a stopwatch. He performs 10 measurements and finds the following values.

Measurement	Value(s), x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	4.07	-0.179	0.03204
2	4.59	0.341	0.11628
3	4.32	0.071	0.00504
4	3.92	-0.329	0.10824
5	4.29	0.041	0.00168
6	4.45	0.201	0.0404
7	4.12	-0.129	0.01664
8	4.48	0.231	0.05336
9	4.06	-0.189	0.03572
10	4.19	-0.059	0.00348
	$\bar{x} = 4.249$		$\sum_{i=1}^{10}(x_i - \bar{x})^2 = 0.38085$

Using the definitions above we find for the standard deviation:

$$\sigma = \sqrt{\frac{1}{10-1} \cdot 0.38085} = 0.2057$$

This implies that for 7 out of the 10 measurements the deviation from the average is smaller than 0.2057.

The accuracy of the measurement is now given by:

$$\delta x = \frac{0.2057}{\sqrt{10}} = 0.065048$$

The final result is: $\boldsymbol{x = 4.25 \pm 0.07\ s}$

If the student had only performed the first 5 measurements he would have found the following result (check this yourself):

$$x = 4.24 \pm 0.11\ s$$

Which is similar to that obtained from 10 measurements, except that the uncertainty is larger by a factor of $\sqrt{2} \approx 1.4$.

COMBINING RESULTS

If we have to determine a quantity that is a function of several independent experimental parameters how should we than proceed? Statistical theory tells us that the best value for this quantity is given by using the best values for the experimental parameters. If we write the quantity *S* as a function of *x,y,z,….* i.e. *S = f(x,y,z,…)* then the best value for S is simply given by:

$$S_{best} = f(x_{best}, y_{best}, z_{best}, ...) = f(\bar{x}, \bar{y}, \bar{z}, ...)$$

The situation for the errors is more complicated and is given by the following expression.

$$(\delta S)^2 = (\frac{\partial f}{\partial x})^2 \delta x^2 + (\frac{\partial f}{\partial y})^2 \delta y^2 + (\frac{\partial f}{\partial x})^2 \delta z^2 + ...$$

As can be seen from the table below this expression often simplifies dramatically.

Function	Error
$S = x + y$	$\delta S = \sqrt{(\delta x)^2 + (\delta y)^2}$
$S = x \cdot y$	$\frac{\delta S}{S} = \sqrt{(\frac{\delta x}{x})^2 + (\frac{\delta y}{y})^2}$
$S = \ln x$	$\delta S = \frac{\delta x}{x}$

Now lets return to our example. The vibrational period, *T*, of a object suspended from a spring is related to the mass, *m*, of the object and the spring constant, *k*, via:

$$T = 2\pi\sqrt{m/k}$$

When we know the mass of the object we can determine the spring constant by rewriting the above expression:

$$k = \frac{4\pi^2 m}{T^2}$$

The student has found before: $T = 4.25 \pm 0.07$ s

Using a balance, resolution 1 gram, he finds for the mass: $m = 0.1385 \pm 0.0005$ kg

The best value for the spring constant is than given by:

$$k = 0.29506 \text{ Nm}^{-1}$$

The uncertainty in this value is determined by the uncertainty in the period, *T*, which is about 1.65% and the uncertainty in the mass, *m*, which is 0.36%.

After some mathematics one finds for the uncertainty δk the following expression (derive yourself):

$$\frac{\delta k}{k} = \sqrt{(\frac{\delta m}{m})^2 + 4(\frac{\delta T}{T})^2}$$

Substitution of the above mentioned errors gives:

$$\frac{\delta k}{k} = \sqrt{0.36^2 + 4 \cdot 1.65^2} = 3.32\%$$

thus δk=3.32% × 0.29506 = 0.0098

As final result we find:

$$k = 0.295 \pm 0.010 \text{ Nm}^{-1}$$

Note that the error is mainly determined by the error in the time measurements.

LINEAR REGRESSION

Often the results of measurements are to be compared to a known function and the constants for this function, which often are the physical or chemical important properties, are to be determined from the data. In the example above the student could measure the vibrational period of the spring as a function of the suspended mass and from this derive the spring constant. Besides obtaining a more accurate value for the spring constant this way (there are more measurements), also the validity of the expression given above is tested.

Often the relation between two variables is linear:

$$y = a + bx$$

or has a functional form that can be linearized:

$$y = a\,e^{bx} \qquad \Rightarrow \qquad \ln(y) = \ln(a) + bx$$
$$y = bx^2 \qquad \Rightarrow \qquad \sqrt{y} = \sqrt{b}\;x$$

In this case the constants can be derived using a method known as linear regression provided the following criteria are met:

- The independent variable, *x*, is precisely known. Usually it is sufficient that the error in x is much smaller than in y, as was the case for the above given example.

- The error in the dependent variable, y, is independent of x.

If we have sets of data points(x_1, y_1), (x_2, y_2),….(x_N, y_N) then the constants *a* and *b* are given by the following relations:

$$a = \frac{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i \cdot y_i}{\Delta}$$

$$b = \frac{N\sum_{i=1}^{N} x_i \cdot y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\Delta}$$

where $$\Delta = N\sum_{i=1}^{N} (x_i)^2 - \left(\sum_{i=1}^{N} x_i\right)^2$$

The uncertainties in these values are given by:

$$\delta a = \sigma_y \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{\Delta}}$$

$$\delta b = \sigma_y \sqrt{\frac{N}{\Delta}}$$

where

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{N} (y_i - a - bx_i)^2}{N-2}}$$

We now have to answer the question whether the data points are well represented by the given function. To do so we have to look at the correlation coefficient, which is defined as:

$$r = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}}$$

When the value of $|r|$ is close to 1, *i.e.* $|r|$>0.98, then the data points are well described by the given function.

Now lets return to our example. The student has found the following values for the vibrational period as function of the suspended mass.

Mass, *m* (kg)	Vibrational period *T* (s)	T^2 (s^2)
0.0755	3.04	9.24
0.1385	4.25	18.06
0.2250	5.46	29.81
0.3025	6.32	39.94
0.4100	7.22	52.13

We can rewrite the following equation:

$$T = 2\pi\sqrt{m/k} \qquad \Rightarrow \qquad T^2 = 4\pi^2 \, m/k \qquad \Rightarrow \qquad y = bx$$

With y = T^2 and b = $4\pi^2/k$ and $x = m$

Using the equations given above we find:

$$b = 128.7 \pm 3.3$$

After conversion we find as final result:

$$\boldsymbol{k = 0.307 \pm 0.008 \ \mathrm{Nm^{-1}}}$$

Furthermore we find $|r|$ =0.999 indicating that the functional form used is correct.

GRAPHICAL REPRESENTATION OF DATA

If you expect a certain functional form it is a good idea to plot the data points as you perform your experiment so that you can see right away if something is wrong with a certain measurement.

In case of our example one would plot the square of the vibrational period as function of the mass.