Measurements and Uncertainties Exercises

Joanna Haigh

Contents

Measurements and Uncertainties Practical Exercises

1. Introduction

The purpose of this session is to help you to gain experience of the correct treatment of experimental uncertainties (errors) in the laboratory. Some of the exercises in this session will make use of your calculator, as this is what you are likely to use while you are working in the laboratory. Some of the exercises also will make use of *Origin Pro* for the calculations. This software package is similar to Excel but its provisions of mathematical analysis and graphing make it a better choice for scientific (rather than business) applications. This three-hour session follows on from the introductory lecture on Measurements and Errors and is designed to reinforce the ideas introduced in the lecture with straightforward practical examples. If you are unsure about anything in this document, please ask a demonstrator – they are there to help you.

Most of the mathematics used for data analysis and error handling is fairly straightforward, but it is important to practise applying these ideas. Data analysis and error handling are practical skills which are used constantly in the laboratory. Practical work is not just about collecting good data but also about the correct analysis of the data and extraction of results from it. This includes the determination of a final error – which tells you how reliable the results are and how they can be used.

When you are working in the laboratory, and presenting your results in laboratory reports, you will always be expected to treat experimental uncertainties properly and to show how you have arrived at your final quoted range of values. The exercises you will go through in this session should introduce you to all the ideas that you will need in your first year laboratory work for the presentation and analysis of data.

When working in the laboratory, you will often need to use your calculator for simple calculations. (We recommend Casio fx-83ES.) For example, you will need to check that your measurements make sense and are giving roughly the expected results. Or you may need to estimate the uncertainty of a set of measurements to check how accurately you are able to make the measurements. At a later stage you will analyse your measurements using an *Origin* spreadsheet in order to make further calculations (e.g. calculating means and standard deviations of sets of data), and to present the data in tables, plot graphs, and extract slopes and intercepts together with their errors.

Important points to remember

1. A measurement without units is meaningless.
2. A measurement without an estimated range of uncertainty (or *error*) is also meaningless.
3. Always quote the final result of a measurement in the form $a \pm b$ with a (measurement) and b (error) given to the same number of decimal places. For example, a measurement of the speed of light might give: $c = (2.94 \pm 0.04) \times 10^8$ m s^{-1}. This expression means that the true value is likely (the probability being approximately 68%, see below) to lie between 2.90×10^8 m s^{-1} and 2.98×10^8 m s^{-1},. i.e. if you repeat the measurement a large number of times about 68% (i.e. two-thirds) of the results will lie between 2.90×10^8 m s^{-1} and 2.98×10^8 m s^{-1}.
4. Errors are usually only known roughly, so quote your errors to no more than two significant figures. One significant figure is usually sufficient. (i.e. 2.94 ± 0.04, or possibly 2.942 ± 0.038, but not 2.94217 ± 0.03794).

2. Means, Standard Deviations, and Standard Errors

Often in an experiment you will make several measurements of a quantity. By taking the average of many measurements, the effect of *random* errors can be reduced because errors of opposite signs tend to cancel each other. Furthermore, the precision of the final result may be estimated by looking at the spread of the data.

NB taking multiple measurements cannot give information on any *systematic* errors which may be present.

When a very large number of measurements are made the spread of results about the mean value, if determined by random fluctuations, will be represented by a *Normal Distribution* (also called the *Gaussian distribution*). This gives the probability $P(x)dx$ of an individual measurement lying between *x* and $x + dx$. The distribution plotted in Fig. 1 has a *mean* of $\bar{x} = 0$, and a *standard deviation* of $\sigma = 1$, although this will not be the case generally. It is always the case, however, that it is normalised so that the total area under the curve is unity (i.e. a measurement is certain to lie somewhere).

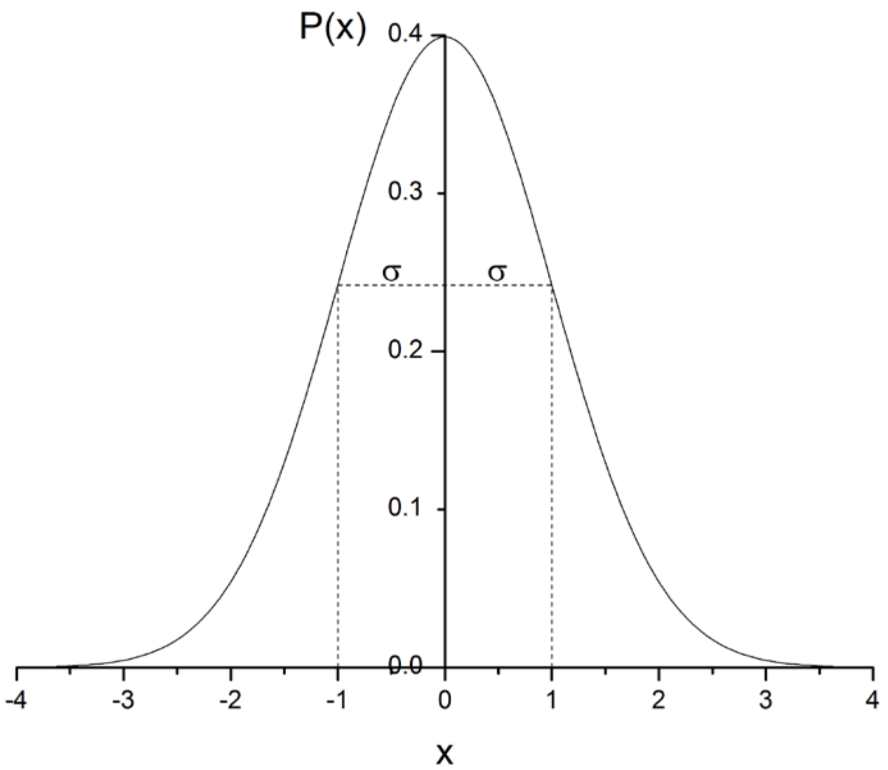


Figure 1 The Normal Distribution. The general form is $\boldsymbol{P(x)} = \frac{\mathbf{1}}{\sigma\sqrt{\mathbf{2}\pi}}\boldsymbol{exp}\left[-\frac{(x-\bar{x})^2}{2\sigma^2}\right]$. In this example $\bar{\boldsymbol{x}} = \mathbf{0}$ and σ=1.

Two important values of the Normal Distribution:

about 68% of the measurements are within ± 1 standard deviation of the mean

about 95% of the measurements are within ± 2 standard deviations of the mean.

Example: Here is a set of values obtained for the wavelength of a sound wave:

0.76m	0.79m	0.84m	0.75m	0.80m	0.79m

We write these as x_1, x_2, x_3, x_4, x_5, and x_6 and the whole set can be referred to as x_i with the suffix i = 1,2,…,N where the number of measurements made N = 6 in this case.

There are three important quantities that we can evaluate from this set of measurements: the mean, the sample standard deviation and the standard error of the mean.

2.1. Mean.

This is the average of the measurements, and is the best estimate of the true value of the quantity that is being measured. It is written as $\bar{x}$ and is calculated from

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

Exercise 1. Using your calculator you can find this *either* by adding the numbers together and dividing by 6, *or* by using the calculator's built-in statistical functions to calculate the mean directly. You **must** know how to use both ways reliably.
[Find online the manual for your calculator now if you don't know how to use its statistical functions.]
Whichever way you calculate it you should find the value 0.788m. Note that the data do not justify any more precision than is implied by giving 3 decimal places so we do not, for example, write the result as 0.788333m.

[NB Answers to Exercises are given at the end of this document.]

2.2. Sample Standard Deviation.

The standard deviation σ is a measure of the spread of the measurements and is defined as the square root of the mean square deviation. For a sample of N measurements the best estimate of σ is given by s (sometimes written as $\sigma_{n\text{-}1}$) in the formula

$$s^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} x_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2\right]$$

[See lecture for derivation].

Again, you can calculate s manually using your calculator, but it is much easier to calculate it using the built-in statistical functions.

Note that unless the number of measurements is rather large the range of values of s that come out of separate sets of measurements is wider than you would expect, so s is never very accurately determined in real measurements (see Exercise 3 below for examples of this). As we will see, the exact size of the errors are never very well known, so rough calculations with error values are generally acceptable.

Use your calculator to find out the value of s for the data above. You should find that s (or $\sigma_{n\text{-}1}$) = 0.03m.

2.3. Standard Error of the Mean.

We have calculated the mean $\bar{x}$ which is an estimate of the true value of the quantity we are measuring. How accurate is this estimate? The sample standard deviation s is *not* a measure of the error in this estimate. We need the Standard Error of the Mean, written as σ_m, which tells you the accuracy with which the mean of the data points gives the true value of the quantity you are measuring (which we will call x_0).

As you take more and more measurements, the standard deviation of the measurements stays the same (the behaviour of the apparatus is not changing) but the accuracy with which they determine the true value of x_0 improves. This is because combining more measurements has the effect of reducing random errors by averaging.

The formula for σ_m is

$$\sigma_m = \frac{s}{\sqrt{N}}$$

σ_m falls as $1/\sqrt{N}$ compared to the sample standard deviation s so, for example, in order to double the accuracy of your estimate for x_0, you have to take 4 times as many data points [again remember that this does not affect any systematic errors].

[*An aside for the expert*: even if the sample data do not follow a Normal Distribution the Central Limit Theorem tells us that the statistics of many samples will follow a Normal Distribution with σ_m as the standard deviation: more on this in the second-year course on Statistics and Probability.]

Calculate σ_m using this formula for the data above. You should find the value 0.013m.

Summarising the results above, for the data set given, you would quote the following for your best estimate of the true value of the wavelength and its error:

$$\lambda = 0.788 \pm 0.013 \text{ m.}$$

Quoting the standard error is a more quantitative way to describe the accuracy than simply using an appropriate number of decimal places. It would also be acceptable to write this as:

$$\lambda = 0.79 \pm 0.01 \text{ m.}$$

Note that this value is quoted with limits of $\pm\,\sigma_m$. This means that the true value is likely (68%) to lie within the range $\lambda - \sigma_m$ to $\lambda + \sigma_m$. However, about 33% (one-third) of the time it will lie outside this range. Sometimes errors are quoted with the value $\pm\,2\sigma_m$ when the author wants to give a limit on the range of possible values rather than an indication of the likely error. This is sometimes referred to as “95% confidence level” as for large N, 95% of all measurements lie within $\pm\,2\sigma$ of the mean. If you do this, you should always state clearly that the uncertainty quoted is 2 standard errors.

Exercise 2. The following measurements are obtained in an experiment to measure the speed of light (all in units of 10^8 ms^{-1}):

2.95	3.04	2.98	3.10	3.01
2.89	2.96	2.47	3.02	2.97

Find the values of $\bar{x}$, s and σ_m for this set of data. How would you quote the final result? Is it consistent with the true value of the speed of light?

Now look at the data more carefully, and you should notice that one of the data points is a long way away from the others. This might be an indication that a mistake was made in taking or writing down this data point. Because this data point is well outside the range of the others (it's about 3 standard deviations away), you might be justified in discarding this point and redoing the calculation. How does that affect the results?

2.4. Use of Origin

In laboratory experiments you will often need to plot graphs of your results and for this we use *Origin*. This software also provides a multitude of data analysis tools. Here you will need to use a relatively small number of its features but it is advised that you familiarise yourself with its features as many will be required in your lab work during the year. A short introductory guide to *Origin* is provided in your lab. handbook.

Start *Origin* and enter the data from Exercise 2 into a column. Put appropriate text into the Long Name and Units boxes. Statistics on the data in a column can be calculated as follows:

Select the column
Click on Statistics in the menu bar then
Descriptive Statistics
Statistics on Columns
Open Dialog

A window will open with many options. You should see in the Data Range box a reference to the column you have chosen. In the Quantities to Compute/Moments select as many parameters as you find interesting but include Mean, Standard Deviation and SE of Mean. Click OK.
If you have used the default Output Settings a new sheet (called DescStatsOnCols 1) will appear. On it you will find a Table containing the values requested. Check that the values are the same as you found using your calculator [**NB Standard Deviation in *Origin* gives the sample value s not the population value σ**].
Repeat the calculation with different settings, this is easily accomplished by clicking the small padlock icon on the results sheet and selecting Change Parameters.

For the rest of the exercises here you will need to load the *Origin* files Exercise3 etc which can be found on Blackboard (http://learn.imperial.ac.uk) under the Year 1 Laboratory and Computing, Measurements and Uncertainty section.

Exercise 3. In the data file you will find five sets of data, each having 20 points. These numbers were generated using a random number generator and have a normal distribution with a fixed mean and standard deviation.
Calculate the mean and standard deviation of each of the samples separately, this is easily achieved by selecting all five columns on the data sheet and when choosing parameters for the Statistics on Columns, make sure that Independent Columns is chosen in the Input Data section. Do the means of each set come out close to each other? Are the standard deviations of each set the same? What is the standard error of the mean of each set? Are the differences between the means of all the sets similar to

the values of the calculated standard errors? Think hard about the results that you get and whether they correspond to what you expect to see.

Finally, find the value for the mean, standard deviation and standard error of the complete data set. To do this select all five columns on the data sheet and select Combined as Single Dataset rather than Independent Columns.

How would you quote a final result for this complete set of measurements? You will see that the standard error here is much less than the standard deviation of the data points. Note the benefit of using a relatively large number of data points. However, in real experiments you have to consider two other factors in deciding how many data points to take:

1. Is the effort involved in taking lots of data points worthwhile given that the relative gain gets slower as more points are taken?
2. Are there other sources of error (either random or systematic) in the experiment that will then dominate the final result so that there is no point in further reducing the random error on the measurement of this particular quantity?

When you have got to this point, discuss your conclusions with a demonstrator, to check your understanding of the material so far, before moving on.

3. Propagation of Errors

In many cases it is necessary to find the value of a *function* of the quantity actually measured. For example, if you measure the radius r of a circular object but want to know the area (call it *A*), then you need to calculate the quantity $A = \pi r^2$. We therefore need to know the error in *A* given the error in r.

There is a general rule which covers all these cases: the error σ_z in z when $z(x)$ is a function of x is given, for small errors, by:

$$\sigma_z = \left|\frac{dz}{dx}\right|_{x=\bar{x}} \sigma_x$$

where σ_x is the error in x and the gradient is evaluated where $x = \bar{x}$.

3.1. Multiplying Factors.

Exercise 4. We take the function $z = 2x$ as a simple example. Using the column of values in the *Origin* file *Exercise4*, find the mean and standard deviation of the values given for x. Now use Set Column Values (as described in the Performing calculations on data section of Introduction to *Origin* document) to set up a column containing the values of $2x$ and find the mean and standard deviation of these values. Are your results consistent with the general expression above?.
Note that the same formula holds for the standard error (σ_m) as well as the standard deviation, as the factor of $1/\sqrt{N}$ is the same in both cases.

We write the error in a quantity x simply as σ_x. Note that in this context the source of the error estimate is irrelevant: it may refer to a standard (random) error derived from a set of measurements, or it may refer to an estimate for the systematic error in x. The mathematics for the propagation and combination of errors is the same in both cases.

3.2. Powers.

Now consider the case of $z = x^2$. Will the mean of z be the same as the square of the mean of *x?* What do you predict for the relationship between σ_z and σ_x?

Make a column of the values of x^2 and then find the mean and standard error. Are the results what you predicted?

Using the general expression above we can write (approximately, for small σ) in this case:

$$\left(\frac{\sigma_z}{\bar{z}}\right) = 2\left(\frac{\sigma_x}{\bar{x}}\right)$$

which shows that the *fractional* error in *z* is equal to *m* times the *fractional* error in *x*. We will see later that fractional errors (*i.e.* percentage errors) are often useful in error calculations. Check that this formula gives the same result as found above.

Exercise 5. You are now in a position to find the error in a slightly more complicated case by combining the ideas earlier in this section:

Consider how the error in an estimate of the volume *V* of a sphere depends on the error in measurement of its radius *r*. [Remember $V = 4\pi r^3/3$]. If the measurements of the radius have a mean of 24.7cm and a standard error of 0.8 cm. What are the best values of *V* and its error?

4. Combinations of Errors

Another important topic to consider is how to evaluate the error in something that depends on the measured values of two different quantities. For example, to get the length of an object you might measure the position of one end of it on a scale and then measure the position of the other end on the same scale. The length is the difference of the two values, but what is the error in the length, given the errors in the position of the two ends?

There is a general rule for combining errors. If z is a function of x and y the error σ_z in $z(x, y)$ is given by:

$$\sigma_z = \left(\left(\frac{\partial z}{\partial x}\right)^2_{\substack{x=\bar{x}\\ y=\bar{y}}} {\sigma_x}^2 + \left(\frac{\partial z}{\partial y}\right)^2_{\substack{x=\bar{x}\\ y=\bar{y}}} {\sigma_y}^2\right)^{1/2}$$

provided that *x* and *y* are independent quantities in the sense that the random errors in measuring them are not correlated with each other.
NB The curly d in e.g. $\frac{\partial z}{\partial x}$ implies partial differentiation. This means you differentiate the function *z* with respect to *x* while keeping all other variables (*y* in his case) constant.

4.1. Sums and Differences.

Supposing $z = x + y$; show that the formula above would suggest:

$${\sigma_z}^2 = \sigma_x^2 + \sigma_y^2$$

This is called “adding the errors in quadrature”. Note that the *same* equation for σ_z applies in the case $z = x – y$ as with $z = x + y$.
[NB the extension to three or more variables is straightforward.]

Exercise 6. An example is given in the *Origin* sheet Exercise6 which gives a set of values for a quantity x and a separate set of values for a quantity y.

Introduce new columns for $x + y$ and $x - y$ into the table and then find the means, standard deviations and standard errors for x, y, $x + y$ and $x - y$.

Are your results consistent with the theory above?

It is sometimes the case that one error is significantly larger than the other, and if this is true we find that the final error is virtually equal to the largest contributing error. For example, if $\sigma_y = 0.3\sigma_x$, then for $z = x \pm y$, $\sigma_z \approx 1.04\ \sigma_x$. In this case we are justified in ignoring the effect of any error in *y* contributing to the final error in z. It is important to bear this in mind – don't bother doing calculations that are unnecessary because the errors are dominated by one source.

Another important situation where errors are added in quadrature is when a set of measurements have both systematic and random errors.

4.2. Products.

The next case to consider is when $z = xy$. Show that in this case the general formula gives the approximate result:

$$\left(\frac{\sigma_z}{\bar{z}}\right)^2 = \left(\frac{\sigma_x}{\bar{x}}\right)^2 + \left(\frac{\sigma_y}{\bar{y}}\right)^2$$

In other words, the *fractional* errors add in quadrature.
[Again the extension to three or more terms is straightforward].

Note that the same formula applies when $z = x/y$ as for $z = xy$.

Find the standard deviations and standard errors for $z = xy$ and $z = x/y$ using the data in the table and see how your results compare with those given by the above formula.

Again, if one fractional error is much larger than the other, then the final error is just determined by the larger of the two. It is often useful to express fractional errors as a percentage, and then if one error is more than about 3 times the other, you can assume that the final fractional error is just equal to the larger of the two contributing errors.

Exercise 7. The refractive index of a material determines the ratio of the angle of incidence θ_i to the angle of refraction θ_r of light as it crosses the material boundary [Snell's Law $n = \sin\theta_i / \sin\theta_r$]. Use the General Rule above to find an expression for $\frac{{\sigma_n}^2}{n^2}$ in terms of the means and errors in θ_i and θ_r.
If measurements of θ_i and θ_r have means of π/6 and π/8 respectively and the standard error of each is π/200 what are the best values of *n* and its error?

Discuss your conclusions for Exercises 6 & 7 with a demonstrator before moving on.

5. Plotting graphs and line-fitting

The aim of this section is to develop good practice in presenting experimental results as graphs.

5.1. Plotting

Exercise 8. You need the *Origin* file Exercise8 from Blackboard. The first column is temperature (in degrees Kelvin) and the second and third columns the resistivity[1] of copper and aluminium, respectively, measured at each temperature. Select the first two columns of data then use Plot, Symbol, Scatter to produce a graph of the copper values.

Following the instructions in the Introduction to *Origin* document now add error bars of 1K in temperature and 5% in resistivity.

Double click on a data point to edit style, size and colour of the points.

In the top left hand corner of the plot you will see a small grey box with the number 1. This indicates that this plot only has one layer. Double-clicking on this produces a box indicating the data that have been used; changes to the input data can be made by choosing different parameters in the box. Plots may have more than one layer (up to a maximum of 121); this is useful, for example, in cases where different datasets (with different units/scale) need to be plotted on one graph.

5.2. Straight line fits

Often with physics experiments we want to test a hypothesis that predicts the data will behave in a way that conforms to a mathematical model. A simple model might predict that two variables have a linear relationship, so that a plot of one against the other would be a straight line. In all cases we can test the hypothesis by seeing how closely the actual measurements match the predicted line.

Now add a straight line fit to your resistivity plot using Analysis, Fitting, Linear Fit.

In the Dialog Box there are lots of options for the fitting and for output.

A Table will appear on the plot giving various measures of the fit, as selected in the options. In the top left corner, below the grey layer number box, will appear a small padlock symbol: clicking on this will reveal a menu including the option to Change Parameters in the fitting options which is very helpful for quick adjustments; try some out.

As well as the summary statistics box on the plot the fitting procedure can produce a lot more information on a different worksheet. An example, for the Cu resistivity data, is given in Figure 2.

Origin gives results to 5 or more significant figures. It doesn't recognise that when, for example, we write a value $1.76\text{x}10^{-8}$ Ω m for resistivity we don't measure it to be $1.760000\text{x}10^{-8}$ Ω m. When you use the results you should generally quote the error to not more than 1 or 2 significant figures, and the main result to the same number of

[1] The resistivity ρ of an electrically-conducting material is defined as the constant of proportionality in the relationship between the resistance R of a wire made of the material, its length L and its cross-sectional area A : $R=\rho L/A$. Its unit is the ohm-meter (Ω m).

decimal places. Thus from Figure 2 we would deduce that the intercept of the resistivity curve is at (-2.31±0.75) x 10^{-9} Ω m.

1 Linear Fit (06/09/2011 17:36:14)

Notes

Input Data

	Input X Data Source	Input Y Data Source	Range
	[Book1]Sheet1!Temperature	[Book1]Sheet1!Resistivity Cu	[1*:9*]

Masked Data - Values Excluded from Computations

Bad Data (missing values) -- Values that are invalid and thus not used in computations

Parameters

		Value	Standard Error	95% LCL	95% UCL
Resistivity Cu	Intercept	-2.31034E-9	7.45865E-10	-4.07403E-9	-5.46656E-10
	Slope	6.59667E-11	2.61962E-12	5.97722E-11	7.21611E-11

Statistics

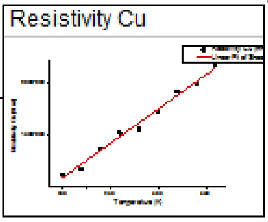
	Resistivity Cu
Number of Points	9
Degrees of Freedom	7
Residual Sum of Squares	1.15289E-18
Pearson's r	0.99453
R-Square(COD)	0.98908
Adj. R-Square	0.98752
Root-MSE (SD)	4.05831E-10
Norm of Residuals	1.07373E-9

Summary

ANOVA

		DF	Sum of Squares	Mean Square	F Value	Prob>F
Resistivity Cu	Model	1	1.04438E-16	1.04438E-16	634.11921	3.97504E-8
	Error	7	1.15289E-18	1.64698E-19		
	Total	8	1.05591E-16			

Fitted Curves Plot



Residual vs. Independent Plot

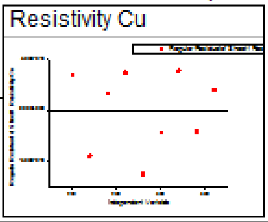


Figure 2 Results from simple linear regression of Cu resistivity data.

Some of the most useful information that can be derived/output are:

- Confirmation of the Input Data used in the calculation.
- The number of degrees of freedom *DF* in the calculation[2]
- The values derived for the slope *m* and intercept *c* of the line, along with the standard errors of those values and the 95% confidence limits[3].

[2] This is the number of data points *n* minus the number of parameters *p* to be fitted, which in this case is 2 *viz m* and *c*.

- The correlation *r* between the x and y values.
- The Residual Sum of Squares *RSS* which is the sum of squares of the differences between the points and the fit line (see also the bottom plot in Figure 2)
- Root Mean Square Error $= \sqrt{RSS/DF}$ (of the fit).
- Total Sum of Squares *TSS* which is the sum of squares of the differences between the points and the mean value (determines the standard deviation of the data, not the fit).
- The R^2 value (coefficient of determination) $R^2 = \frac{Explained\ variation}{Total\ variation} = \frac{(TSS-RSS)}{TSS} = 1 - \frac{RSS}{TSS}$.
  Its value usually lies between 0 and 1, where 1 indicates a perfect fit and 0 indicates that the data is so far from conforming to a straight line that none can be determined. Note that a high value of R^2 may indicate a low value of random error but it says nothing about the presence of systematic error.

Take a look at these values in your output and make sure you understand how they relate to the data and fit.

There is an option to force the line to go through a particular intercept value (Fix Intercept) but think carefully before using this. Supposing theory says a line should go through the *Origin*: the presence of a systematic error in the measurements will be revealed by the line missing the *Origin*. Not only that but by forcing the line to go through the *Origin* you would make it have the wrong slope. Try this out on the resistivity data and see how the values of *m* and *c* respond.

5.3. Multiple datasets

You can add a second (or subsequent) curve to an existing plot. Return to the data sheet and add a column giving 5% for the error in the Aluminium values. Select the 4 columns necessary to plot the Al data with error bars. Now return to the Graph sheet and choose Graph, Add Plot to Layer, Scatter. Edit symbol size etc as for copper data.

Remove the exponent values (E-08) on the y axis by double-clicking the y axis and choosing Tick Labels, Divide by Factor and choosing an appropriate value. Edit the y axis title correspondingly (double-click on its box and use the Format Toolbar. Another useful facility is the Symbol Map which you can open by right-clicking on the box once opened).

You may also need to edit the legend which you can do either by double-clicking its box or using Graph, New Legend.

You should end up with a plot looking something like **Figure 3**.

---

[3] For a large number of degrees of freedom (i.e. data points) a range of ± 1 standard error is 68% likely to contain the correct value and ± 2 standard errors is 95%. These percentages are reduced for a smaller number of degrees of freedom.

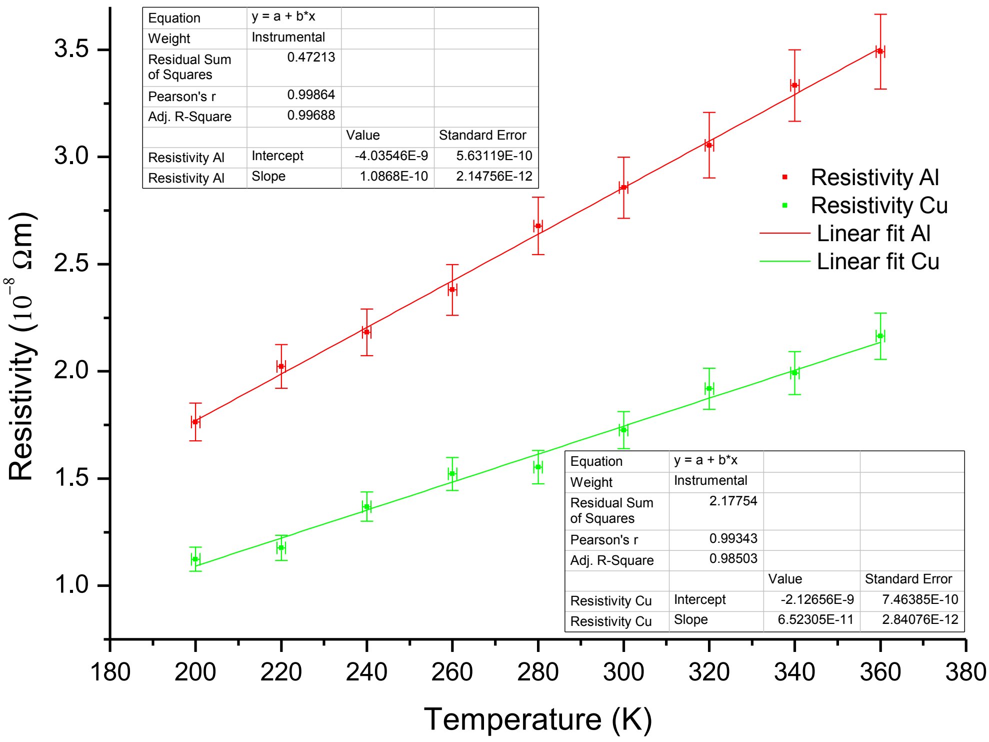


Figure 3 ***Origin*** **plot of resistivity as a function of temperature for copper and aluminium with error bars, linear fits and derived statistics.**

6. Practical Example

Imagine an experiment where the resistance of a pair of identical resistors is to be found. A measurement is made of the voltage difference across the two resistors and the current running through them is also measured. We expect the resistance of each resistor to be described the equation:

$$R = \frac{1}{2}\frac{V_1 - V_2}{I}$$

where V_1 and V_2 are the voltages at the two ends of the resistors and I is the current through them.

Exercise 9

Consider two approaches to finding the value of R.:

1. We take one measurement of each of V_1, V_2 and I and accept the equipment manufacturer's error estimates giving the following values: $V_1 = 6.9 \pm 0.5$ V, $V_2 = 0.7 \pm 0.1$ V and $I = 0.43 \pm 0.03$ A. Find a value for R and its error σ_R using the appropriate methods for combining errors.
2. We take a series of measurements of V_1, V_2 and I with results as given in the *Origin* file Exercise9. Plot (V_1 - V_2) against I and use a linear fit to find R and σ_R.

Do the two approaches give the same results?

7. Non-linear fits

So far we have only considered fitting straight lines to a dataset. More complex relationships might require the line to be represented by e.g. a polynomial or an exponential function. These, and other, fits can also be achieved using other options under Analysis and Fitting.

Exercise 10. Read in the data[4] from the *Origin* file Exercise10.
Try fitting a polynomial of order 2.
Add y-errors of magnitude 10% before mid-1993 and 2% after that; how does that affect your fit?
Experiment with other functions. Given that the incidence of cosmic rays is modulated by the 11-year solar activity cycle try a plausible sine function (use Parameters to fix the period).

8. Histograms

It is often useful to look at spreads in measurements using histograms.
Exercise 11 Using the dataset in Exercise11[5] in *Origin* use Plot, Statistics, Histogram to plot a quick result. Now try varying the bin sizes: select data column then use Statistics, Descriptive, Frequency Counts, Computation Control. The results will appear in a new sheet, you can then plot the histograms using these data.

---

[4] The first column contains the date and the 2nd and 3rd columns contain indirect measurements of cosmic rays made at ground stations in Colorado, USA and Kiel, Germany. Cosmic rays are energetic particles that are generated in outer space and pass through the Earth's atmosphere. As they collide with atmospheric particles, they disintegrate into smaller pions, muons and the like, producing a *cosmic ray shower*. These particles can be measured on the Earth's surface by neutron monitors. The data are monthly mean neutron counts in units of 100 per hour. The neutron counts at Kiel are consistently higher than those at Colorado because of its more northerly latitude: the charged particles are steered around the Earth's magnetic field lines so that there is a much higher incidence of cosmic rays at the poles than at the equator.

[5] These are real exam results from one of the 3rd Year Comprehensive Papers!

Bibliography

Website

The Measurements and Errors material can be found on the web by going to First Year Lab and Computing on Blackboard.

Textbooks

Young and Freedman, sections 1.1-1.6, give an elementary introduction.

Hughes, I. F. and T. P. A. Hase *Measurements and their Uncertainties* (Oxford University Press, 2010).

This is a friendly slim volume with nice examples and useful exercises (if you want to test your understanding). It contains more than you need for First Year Lab but will also be useful in later years.

Squires, G. L. *Practical Physics* (Cambridge University Press, 4th edition 2001).

This is a good general textbook on experimental measurement and the treatment of errors. Chapters 1 to 5 deal with errors and chapters 6 to 13 deal with experimental technique and record keeping.

For more details of error theory see:

Taylor, J. R. *An Introduction to Error Analysis* (University Science Books, 2nd edition 1997).

SOLUTIONS TO EXERCISES

Exercise 1

Mean = 0.788 m
Sample Standard Deviation (s) = 0.032 m
Standard Error of the Mean (σ_m) = 0.013 m

Exercise 2

Mean = 2.939 $\times 10^8$ ms^{-1} (or 2.94 $\times 10^8$ ms^{-1})
Sample Standard Deviation (s) = 0.174 $\times 10^8$ ms^{-1} (or 0.17 $\times 10^8$ ms^{-1})
Standard Error of the Mean (σ_m) =0.055 $\times 10^8$ ms^{-1} (or 0.06 $\times 10^8$ ms^{-1})

The result should be quoted as (2.94 $\pm$ 0.06) $\times 10^8$ ms^{-1} (this is just consistent with the correct value of c of 3.00 $\times 10^8$ ms^{-1}).

With the unreliable point taken out:
Mean = 2.991 $\times 10^8$ ms^{-1} (or 2.99 $\times 10^8$ ms^{-1})
Sample Standard Deviation (s) = 0.060 $\times 10^8$ ms^{-1} (or 0.06 $\times 10^8$ ms^{-1})
Standard Error of the Mean (σ_m) =0.020 $\times 10^8$ ms^{-1} (or 0.02 $\times 10^8$ ms^{-1})

The result should now be quoted as (2.99 $\pm$ 0.02) $\times 10^8$ ms^{-1} (note the lower final error).

Exercise 3

For the five data sets we obtain:

	SET 1	SET 2	SET 3	SET 4	SET 5
MEAN	6.50	6.64	6.34	6.21	6.50
STD DEV	0.88	0.85	0.65	0.82	0.69
STD ERROR	0.20	0.19	0.14	0.18	0.15

The overall mean is 6.44 and its standard error is 0.08.

Note that, even with as many as 20 data points per set, the standard deviations of each set can be quite different from each other. These data points were generated at random with a mean of 6.5 and a standard deviation of 0.8. You can see that the final results for the total set of 100 points are consistent with these values.

Exercise 4

The results obtained are

	X values	2X values	X^2 values
MEAN	0.396	0.793	0.159
STD DEV	0.040	0.080	0.031
STD ERROR	0.009	0.018	0.007

The calculated standard error for $2x$ is $2 \times 0.009 = 0.018$ (as above).
The calculated standard error for x^2 is $2 \times 0.396 \times 0.009 = 0.007$ (as above).

Exercise 5

V = 63121 ± 6133 cm^3 or better (6.3 ± 0.6) × 10^4 cm^3 or even better (6.3 ± 0.6) × 10^{-2} m^3.

Can do this either by direct calculation using $\sigma_V = 4\pi r^2 \sigma_r$
or using fractional errors:
σ_r/r = (0.8/24.7) = 0.032 (i.e. 3.2%) so σ_V/V = 3×3.2% = 9.6%, which gives the same error as calculated above.

Exercise 6

	X values	Y values	X+Y Values	X-Y Values	X*Y values	X/Y values
MEAN	26.9	20.3	47.2	6.5	544.1	1.4
STD DEV	2.4	5.1	5.1	6.2	138.2	0.4
STD ERROR	0.8	1.6	1.6	1.9	43.7	0.1
%Error	3	8	3	30	8	10

Note that the errors here do not agree exactly with predictions because there are only 10 data points.

Exercise 7

Differentiate the expression for n with respect to θ_i and θ_r and use the results with the general expression for combining errors to show that:

$$\frac{{\sigma_n}^2}{n^2} = [(\cot\theta_i)^2 + (\cot\theta_r)^2]{\sigma_\theta}^2$$

Insert values to obtain n=1.31±0.06.

Exercise 9

V_{12} = 6.2 ± 0.5 volt (notice that this error is completely determined by the V_1 error)
fractional error on V_{12} is 8%
fractional error on I is 7%
so fractional error on R is $\sqrt{(8^2+7^2)}$ = 11%, giving R = 7.2 ± 0.8 Ω.