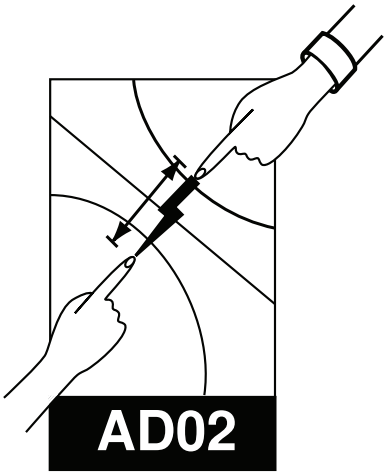
estimating experimental errors



Various August 2011

AD02

1 Introduction

In physics, measurements are almost worthless unless accompanied by an estimate of their validity. Making these estimates is an important part of experimental physics. This document contains the main techniques you will need to use.

In this document, the word **reading** is used to mean a single **measurement** process, e.g. one setting of a micrometer. We reserve the word 'measurement' for the result calculated from a number of readings.

2 Random and systematic errors, accuracy and precision

These concepts are explained in sections 2.1–2.8 below. The expressions which are usually used to evaluate uncertainty are almost always concerned with 'random' or 'statistical' error, although systematic errors are often at least as important and much harder to track down or estimate. In this summary we concentrate on statistical uncertainty because it is amenable to analytical treatment. However, you should always try to identify and quantify systematic errors; it is quite common in experimental physics to find a sophisticated treatment of random errors which gives a misleading apparent accuracy because the systematic errors have been seriously underestimated or even ignored.

2.1 Introduction

As an example of a simple measurement, consider the direct determination of the length of a piece of string. If 12 independent measurements with a ruler having 0.5 mm as its smallest division are made, and the measurements interpolated by eye to 0.1 mm, typical values can be shown as a histogram:

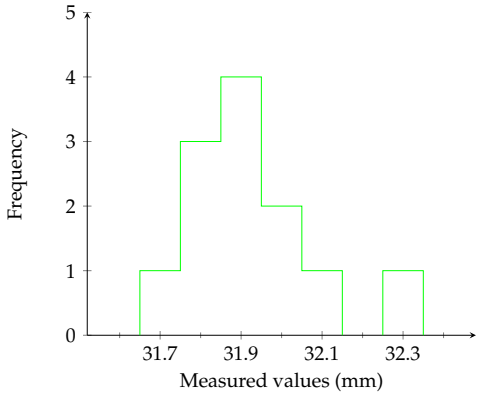


Figure 1: Data obtained when measuring the length of a piece of wire

There are three things to note about this data set:

(i) successive independent measurements do not generally agree in value

(ii) the measurements tend to cluster about some central value

(iii) the clustering shows a statistical regularity that appears in the shape of the histogram or frequency distribution.

We can associate three fundamentally different sorts of errors associated with this, and any other, measurement procedure. These are random errors, systematic errors, and careless errors or mistakes.

2.2 Mistakes and careless errors

The peak in the histogram at 32.3 mm is probably the result of an outright mistake, probably due to careless reading of the scale of the ruler. A measurement which is **known** to contain one or more mistakes should be corrected or discarded. Blunders of this type include:

- misreading of a scale or a measuring instrument; failure to correct for offset zero of measuring scale;
- mistakes in transcribing data e.g. to a computer;
- confusion over units used in the measurement;

Errors of this kind can be eliminated completely by attention to detail; by checking readings and calculations and by exercising care and thought whilst carrying out the experiment and analysing the data.

2.3 Random errors

The spread of the measurements around the central peak of the histogram is the result of random errors in the measurement process. These arise, in this case, mainly from the limitations of the measuring device which prevent the observer obtaining exactly repeatable results. We know that the effect of random errors on the precision of the measurement may be reduced by taking a sufficiently large number of readings. Later we show how from a number n of random trial measurements, we may make a quantitative statement of the reliability of the resulting measurement, when this is expressed as the arithmetic mean of the n separate trials.

2.4 Systematic errors

These are errors which always tend to have the same sign, and thus shift the central peak of the histogram away from the true length of the wire. Examples of systematic errors are those caused by:

- Incorrect or unjustifiably assumed calibration of a measuring scale or instrument.
- Failure to correct for thermal expansion; one-sided illumination of a measuring scale; non-vertical position of a liquid manometer.
- Failure to make necessary corrections for: the effect of atmospheric pressure; variation of gravity with height or latitude; parallax error when taking a reading, etc.

Systematic errors are by their nature more difficult to detect than random errors, but, as they can not be removed by averaging, it is often more important that their magnitudes are estimated and steps taken to prevent them occurring wherever possible.

All experiments include contributions from both random and systematic errors.

2.5 Random errors and precision of measurements

For a measurement expressed as the mean value of the n separate trials, the **precision** is proportional to the reciprocal of the **random** error associated with the measurement. High precision is achieved if the statistical random error is small.

2.6 Systematic errors and accuracy or reproducibility of measurements

Accuracy denotes an experimental result for which the **systematic** error has been shown to be small. Often this claim of high accuracy can be confirmed only by the reproducibility of the results obtained by different experimenters using similar apparatus. A classic situation involving systematic errors is that relative atomic mass determinations made prior to 1920 contained unsuspected systematic errors which were later shown to be ten times the stated experimental errors.

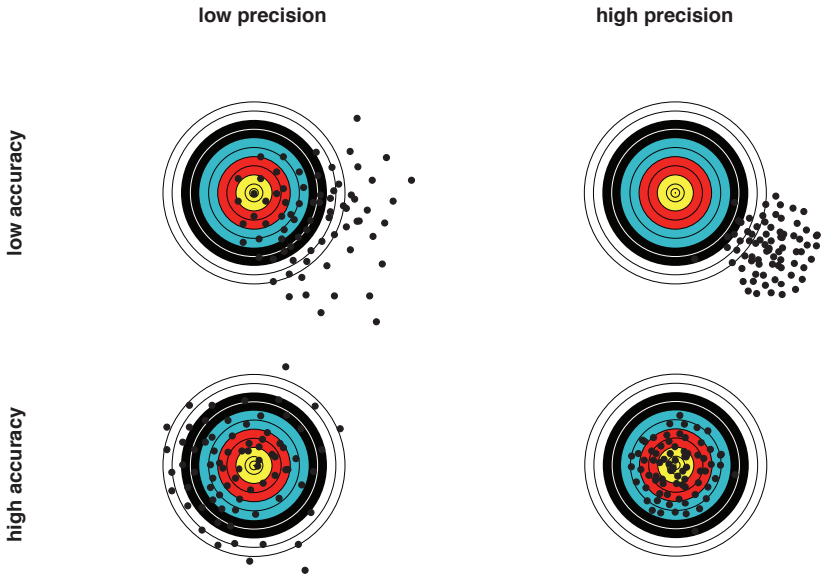


Figure 2: The difference between accuracy and precision.

2.7 Discrepancy

The difference between two measured values reported by two different experimenters, or the difference between the value obtained by an undergraduate and a result published in the peer-reviewed literature is termed a **discrepancy**. This difference is not an error. However, the observer needs to estimate the magnitude of both the random and systematic errors to provide a basis for understanding the origin and interpretation of the discrepancy.

2.8 Engineering tolerance and experimental errors

Engineering tolerances on mechanical and electrical components are often expressed as $X \pm \Delta X$, while experimental measurements of the same parameter may be expressed as $\bar{x} \pm \Delta x$. The tolerance ΔX and the experimental error Δx are not related to one another and they convey different information. The tolerance ΔX gives the range of values, from $X - \Delta X$ to $X + \Delta X$, in which a measured parameter must lie before a component is acceptable for use. Tthe experimental error Δx gives information about the magnitude of the combined random and systematic errors associated with the measurement. In engineering, the measurement techniques used must be such that the error or uncertainty Δx, in the quality control procedures is significantly smaller than the tolerance ΔX set by the manufacturing specifications for those components.

3 Obtaining a result and estimating its accuracy

3.1 The result

Suppose we take N readings of some quantity, for instance, measuring the diameter of a steel rod by taking ten different readings with a micrometer screw gauge. We denote our readings by $x_1, x_2, x_3, \ldots x_N$. Then our best estimate of the result is x_m, the **mean**, or **average**:

$$x_m = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N} \tag{1}$$

This is just the mean of our readings, not, of course, the 'right answer'. Assume that the 'right answer' is the result we would get if we averaged an infinite number of readings. This is intuitively reasonable, as the more readings we take the more our random errors will average out. We call this 'right answer' the **population mean**, denoted by x_μ. It is important to understand that when we take a set of readings, this is a **sample** drawn from the whole **population** of readings which we would obtain if we went on taking readings for ever. Thus, the mean we have actually found, x_m, is just a **sample mean**, that is, it is found from a sample of readings - we could always take more if we wanted to. The **population mean**, on the other hand, is the mean of all possible readings! We cannot ever know what x_μ is, of course - the best we can do is x_m - so it remains to assess how well our result matches the 'right answer'.

3.2 Uncertainty in the result

We thus consider how far x_m is likely to be from x_μ, i.e. we want to make the best estimate of the error associated with our result. First, consider the sample standard deviation s, given by:

$$s^2 = \frac{(x_1 - x_m)^2 + (x_2 + -x_m)^2 + (x_3 - x_m)^2 + \ldots + (x_N - x_m)^2}{N} \tag{2}$$

This definition shows that s^2 is simply the mean square difference of the readings from their average. It is straightforwardly calculated by hand, or, more realistically, using a computer. The value of s is not, of course, the error on our result. For example, we would hope that the error on our result would reduce if we took more readings, whereas it should be clear that if we take, say, 200 readings rather than 100, s would be practically the same in both cases.

The next step involves a further definition. Just as it is useful to consider what x_m would become if we took an infinite number of readings (it would tend towards the 'right answer', x_μ) it is also useful to consider what s would become if we took an infinite number of readings. We call this quantity the **population standard deviation**, σ (Greek letter s). An important property of σ is that the probability of any given reading being within $\pm\sigma$ of x_μ is about 68%. Of course, we can't take an infinite number of readings, so we can't find σ from our limited data any more than we could find x_μ.

However we can make a good estimate of it. In fact, it turns out that our best estimate is given by

$$\sigma^2 = \frac{N}{N-1} s^2 \tag{3}$$

so that provided one has taken enough readings s and σ are very similar. However, if N is small, equation (4.3) shows that s tends to underestimate σ; this can easily be understood by noting that if we have only one reading, s is equal to zero.

The reason for finding σ is that we use it to obtain our estimate of the uncertainty in x_m. For this purpose, we define σ_m, the **standard error on the mean**. It is the probability that x_μ lies within $\pm\sigma_m$ of x_m is 68%. It can be shown that

$$\sigma_m^2 = \frac{\sigma^2}{N} \tag{4}$$

where N is the number of readings we have included in our average. **The standard error on the mean decreases with the square root of the number of readings we take**.

If we now substitute for σ from equation 3, then s from equation 2, we have

$$\sigma_m^2 = \frac{s^2}{(N-1)} \tag{5}$$

$$= \frac{1}{N(N-1)}[(x_1 - x_m)^2 + (x_2 + -x_m)^2 + (x_3 - x_m)^2 + \ldots + (x_N - x_m)^2] \tag{6}$$

This expression is important, as it tells you how to find the uncertainty in your measurement simply using the readings you have taken. It involves $N-1$, because although equation 4 is our basic expression we can't evaluate it directly - we have to go via equation 3 to get a value for σ from our experimentally determined quantity s. However, if using $N-1$ rather than N changes your error estimate significantly, you are 'up to no good', to quote a famous mathematical methods guide, 'Numerical Recipes'. To apply this analysis you need at least five, ideally more readings.

In the case of our ten readings of the diameter of the steel rod, therefore, we would calculate the sample standard deviation s and, making use of equation 5, we would divide s by 3 to get σ_m, the standard error in the mean.

3.3 A note on instrument resolution

Suppose your 10 readings of the diameter of the steel rod were taken with a micrometer which could only be read to an accuracy of 0.01 mm. The readings are as follows, all in mm:

12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25, 12.25

What would be the best estimate of the diameter and its associated error?

As all the readings are the same, the measuring instrument is the limiting factor. The best estimate of the diameter is of course 12.25 mm, but a blind application of the error analysis gives zero for both the standard deviation and the standard error in the mean. All that we can really say with the data we have is that as we never got 12.26 mm or 12.24 mm, the diameter of the rod is closer to 12.25mm than either of these. We could therefore quote ± 0.005 mm as our error estimate. However, note that the accuracy does not improve as we take more readings; we are limited by the resolution (smallest change detectable) of the micrometer.

What if one or two of the readings **had** been different, say 12.26 mm? The error analysis described above is still not reasonable, even though we now do not get an obviously nonsensical result. The variation which is occurring from one reading to another is not a statistical spread, but is still dominated by the resolution of the measuring instrument. If you come across a case of this kind, do not blindly apply formulae, but discuss it with a demonstrator.

3.4 Error analysis - definitions and nomenclature

In practice, the error analysis is carried out using the statistical functions on a computer data analysis package. However, they are written out in full below for completeness, for n readings of a quantity, denoted by x_i, i.e. from x_1 to x_n.

3.4.1 Random error

Deviation between a single measured value and the 'best' value from a set of measurements. Reliability of a determination of the mean of a set of measurements.

3.4.2 Systematic Error

An error that tends to have the same algebraic sign introduced by the process of measurement. It appears in all measurements and shifts the mean value in one direction.

3.4.3 Population distribution

The histogram which would be obtained if all possible (or an infinite number) of trial measurements were to be made.

3.4.4 Trial or sample distribution

The histogram obtained by making a finite number n of measurements of the observed quantity x.

3.4.5 Sample mean m

Definition:

$$\begin{aligned} x_m &= \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i \end{aligned} \tag{7}$$

The mean m estimates the mean of the population distribution μ.

3.4.6 Population mean μ

Definition:

$$\mu = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

For a given number of trials n the best available estimate of μ is given by the sample mean m.

3.4.7 Experimental or sample standard deviation s

Definition:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 \tag{9}$$

s is also called the root mean square deviation. If the measured distribution is Gaussian, with a mean μ and a standard deviation σ,

$$G(x,\sigma) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \tag{10}$$

the probability that the next measurement will fall within the range $\mu \pm \sigma$ is equal to 0.675.

3.4.8 Population standard deviation σ and variance σ^2

If all possible measurements were known, then the population mean μ could be used instead of m to calculate the deviation of each measurement. The variance σ^2 and the standard deviation σ of the population are defined by

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{11}$$

For a set of n measurements, the best approximation to σ^2 is given by

$$\sigma^2 \approx \left(\frac{n}{n-1}\right) s^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - m)^2 \tag{12}$$

It is important to note that the standard deviation (s or σ) determines the width of the histogram or distribution of measurements, and it is not the precision to which the mean (m or μ) is known.

3.4.9 Standard error

For a set of n measurements, the best estimate of σ_m, the standard deviation on the mean, is given by

$$\sigma_m^2 = \sigma^2 / n \tag{13}$$

where σ^2 is the variance of each measurement. σ_m is the standard error of the mean, or more simply, the **standard error**. Although the variance s^2, equation 12, will not change much as more observations are made, since it is an estimate of the variance of the parent distribution, the standard error of the mean, s^2/n, decreases with increasing sample size n. More data help to locate the mean to higher precision.

It is usually good experimental practice to measure a quantity several times and take the average. The precision with which this mean is known is then given by the standard error on the mean, σ_m.

4 Combining measurements

From now on it is assumed that we have found means and standard errors of any quantity we are dealing with. **The symbol σ will be used to refer to the standard error in the mean, and not the population standard deviation**. The first case to consider is: what if we have made two measurements of the same thing, e.g. the diameter of a steel rod? Imagine two different people made the measurements, and they have arrived at different results with different errors. Let the results be x_1 with an error σ_1 and x_2 with an error σ_2. Then the best estimate of the diameter, taking account of both measurements, is

$$\bar{x} = \frac{\frac{x_1}{{\sigma_1}^2} + \frac{x_2}{{\sigma_2}^2}}{\frac{1}{{\sigma_1}^2} + \frac{1}{{\sigma_2}^2}} \tag{14}$$

This expression is simply a weighted mean. Suppose σ_1 is much smaller than σ_2, so that the first measurement is the more accurate. Then $\frac{1}{{\sigma_1}^2}$ is greater than $\frac{1}{{\sigma_2}^2}$; it is reasonable that x_1 should be weighted more strongly than the poorer measurement. To convince yourself that the denominator is necessary, suppose $\frac{1}{{\sigma_1}^2} = 10$ in whatever units we are using, while $\frac{1}{{\sigma_2}^2} = 4$. Then unless we divide by $4 + 10$ our result will be about 14 times larger than either of the two measurements individually.

However, it only makes sense to combine measurements in this way if they are consistent with each other. Thus, if one of our measurements is (1.25 ± 0.01) mm, while the other is (1.45 ± 0.02) mm, then a blind application of our expression gives a final result of 1.29 mm, incompatible with either and obviously spurious. There is clearly a large systematic error in at least one of the measurements we are combining. A simple rule of thumb is that the results are inconsistent if they are much further apart than the sum of their individual standard errors.

We can generalise this method of combining data. Suppose we have n measurements of the same quantity, x_i, each with its standard error σ_i. Then each value is weighted by multiplying it by the factor

$$\omega_i = \frac{\frac{1}{{\sigma_i}^2}}{\frac{1}{{\sigma_1}^2} + \frac{1}{{\sigma_2}^2} + \ldots + \frac{1}{{\sigma_n}^2}}$$

which can be written more compactly:

$$\omega_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_k^2}} \tag{15}$$

To find the weighted mean, we then just find

$$\bar{x} = \sum_{i=1}^{n} \omega_i x_i \tag{16}$$

It may help to look at equation 1 again, which is an unweighted mean. When we take individual readings, we have no reason to suppose that some are more accurate than others, so they are all given the same weight; but when we want to combine the results of different measurements of the same thing (i.e. each calculated from a set of readings), we need a more general expression which enables us to discriminate in favour of the more accurate measurements. Convince yourself that the weights are normalized, that is, that their sum is equal to 1 (use equation 15).

We now consider the standard error in the weighted mean. It must be smaller than any of the individual standard errors, because including more information improves precision. In fact, the result turns out to be σ_w (standing for σ_{weighted}), where

$$\frac{1}{\sigma_w^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \ldots + \frac{1}{\sigma_n^2} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \tag{17}$$

If we combine two measurements of the diameter of the steel rod (as usual, each obtained from a number of readings), and the measurements have the same standard error, then the final error will be smaller than the individual ones by a factor of $\sqrt{2}$. However, if one of the measurements is much more accurate than the other, both the final result and its error will be very close to those of the more accurate measurement.

We now turn to the propagation of errors. This is important, because most experiments try to find the value of a quantity which is derived from measurements of two or more other quantities.

5 Propagation of errors

5.1 Linear dependence

First, consider measuring the length of a rod by finding the two positions on a ruler where the rod's ends lie. We then have x_1 with an error σ_1, and x_2 with an error σ_2 (as ever, we take for granted that a number of readings has been processed to arrive at these results). Our best value of the length of the rod is then $x_2 - x_1$, and the error is given by

$$\sigma = (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}} \tag{18}$$

This time, of course, the error is higher than it is on either of the measurements taken individually. Generalising, we consider a quantity u which is obtained from n quantities we have measured, these being x_1 with an error σ_1, x_2 with an error s, etc., as follows:

$$u = a_1 x_1 \pm a_2 x_2 \pm \ldots \pm a_n x_n \tag{19}$$

(The $\pm$ signs mean that from the point of view of calculating the error in u, it doesn't matter whether the terms on the right-hand side are to be added or subtracted). Then our best estimate of u is obtained by simply substituting our measured results into (14), and the error is given by

$$\sigma(u) = [a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \ldots + a_n^2\sigma_n^2]^{\frac{1}{2}} \tag{20}$$

Adding the squares of quantities like this and then taking the square root of the result is called **adding in quadrature**. The result is less than the ordinary sum, so it would be pessimistic just to add the individual errors. We can see that this is plausible by considering again the case of the rod. There is a 50% chance that the measurements of the two ends are in error in the same sense, i.e. so that they tend to compensate.

5.2 Product or quotient

This introduces a new idea, that of a **fractional error**. For any given measurement, this is the value of σ, expressed as a fraction of the result of the measurement itself. We start with a simple example, the area A of a rectangle, expressed in terms of the lengths of the sides. We assume we have made measurements of these lengths, obtaining x and y, with standard errors σ_x and σ_y respectively. Then the best estimate of the area is simply xy, while the standard error in the area is given by

$$\left(\frac{\sigma_A}{A}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2 \tag{21}$$

Again, we are adding in quadrature, but this time we add the fractional errors in the individual measurements to obtain the fractional error in the product.

We now generalise, so that we deal with any power law, positive, negative or fractional, thus including products and quotients. Suppose we wish to find the error in a quantity u of the general form

$$u = x^a y^b z^c \ldots \tag{22}$$

where x, y, z are independent quantities we have measured. The best value of u itself is obtained by substituting our measured results for x, y, z into equation 22, and the standard error is given by

$$\left(\frac{\sigma_u}{u}\right)^2 = a^2\left(\frac{\sigma_x}{x}\right)^2 + b^2\left(\frac{\sigma_y}{y}\right)^2 \ldots \tag{23}$$

The fact that fractional errors are used here is easier to understand if you remember that the various quantities x, y, z do not necessarily have the same dimensions. The point about equation 23 is that all the terms in the sum are dimensionless.

5.3 Combination of n independent measurements

If a given quantity, e.g. Rydberg constant for atomic hydrogen, is derived from independent measurements of the wavelengths of n spectral lines in the Balmer series, the 'best' estimate of the measured quantity is the weighted mean $\bar{x}$ given by

$$\bar{x} = \sum_{i=1}^{n} w_i x_i \tag{24}$$

where the weights w_i are derived from the standard deviations σ_i associated with each of the independent measurements using

$$w_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \tag{25}$$

The standard error in the mean $\bar{x}$ is given by σ_m where

$$\frac{1}{\sigma_m^2} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \tag{26}$$

5.4 The general case — differential error analysis

We assume that the derived quantity u is related to the J independent directly measured quantities $(x, y, z, \dots)$ by the functional relationship

$$u_i = f(x_i, y_i, z_i, \dots) \qquad \text{and} \qquad \bar{u} = f(\bar{x}, \bar{y}, \bar{z}, \dots) \tag{27}$$

where the bar signifies the mean value and the function f may be additive, multiplicative, exponential or some other combination. The function f is assumed to be continuous and differentiable.

The 'best' value of the observed quantity $\bar{u}$ is the mean value defined by

$$\bar{u} = f(\bar{x}, \bar{y}, \bar{z}, \dots) \tag{28}$$

The variance in u, which results from small changes in the measured values $(x, y, z, \dots)$ can be determined from the calculus of partial derivatives:

$$\delta u = \frac{\partial f}{\partial x_i}\delta x_i + \frac{\partial f}{\partial y_i}\delta y_i + \frac{\partial f}{\partial z_i}\delta z_i + \dots \tag{29}$$

and if we the square and average over the series of measurements (where cross terms like $\delta x_i \delta y_i$ vanish as the measurements are uncorrelated) then the standard deviation $s(\bar{u})$ in the mean $\bar{u}$ (or the standard error in the mean) is more generally given by

$$s(\bar{u}) = \sqrt{\sum_{j=1}^{J} \left(\frac{\partial u}{\partial x_i}\right)^2 s(\bar{x}_j)^2} \tag{30}$$

where $s(\bar{x}_j)$ is the standard deviation in the mean of the jth component property and the partial derivatives are evaluated at $x = \bar{x}$, $y = \bar{y}$, $z = \bar{z}$, etc.

6 Graphs

In many experiments, you will need to draw a graph, and find the error on the slope or intercept, or both. First, read section 1 which deals with establishing a sound experimental procedure. A computer is then used to plot the data, and supply values of the slope and intercept with their standard errors, calculated as described above. However, there are a few points to watch:

(i) Record your results on paper, and then enter them onto the computer, while you are actually doing the experiment. If there is something going wrong, e.g. if there is one point that does not fit on an otherwise straight line, it is a good idea to find out while you can still do something about it. Keeping a paper record ensures data does not get accidentally mis-entered or deleted

(ii) You may be tempted to do a 'weighted best fit' rather than an ordinary unweighted one because the measurements you are going to plot have different standard errors. It may be that the differences aren't significant, particularly if there are only a few readings. So don't weight your results unless you can think of a good reason why some points should be more accurately determined than others.

(iii) Note that computer software often assumes that you only have appreciable errors in the variable you are calling y. This is not always true.

It is easy to do curve fitting to experimental data using a variety of computer packages, but for completeness, the equations for line fitting are described below.

6.1 Unweighted straight line fit

Assume the data consist of n sets of measured quantities (x_i, y_i) and that only the dependent variables y_i are subject to experimental error. If the precision of individual data points is unknown, assume that all the weights are equal, $w_i = 1/n$, and use an unweighted fit to a straight line

$$y = a + bx \tag{31}$$

Best fit parameters

The best fit is taken to be the function which minimises the equally weighted sum S of the squares of the deviations d_i of the observed points from the straight line. Thus we have

$$S = \frac{1}{n}\sum_{i=1}^{n} d_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - a - bx_i)^2 \tag{32}$$

$$\text{subject to the constraints} \qquad \frac{\partial S}{\partial a} = 0 \quad \text{and} \quad \frac{\partial S}{\partial b} = 0 \tag{33}$$

The following results can be derived:

$$\begin{aligned} \text{gradient} \qquad & b = \frac{[\xi\eta]}{[\xi^2]} \\ \text{y-intercept} \qquad & a = \bar{y} - b\bar{x} \\ \text{x-intercept} \qquad & x_{\text{int}} = -\frac{a}{b} \end{aligned} \tag{34}$$

where

$$\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i & \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ \xi_i &= x_i - \bar{x} & \eta_i &= y_i - \bar{y} \\ [\xi^2] &= \frac{1}{n}\sum_{i=1}^{n} \xi_i^2 & [\xi\eta] &= \frac{1}{n}\sum_{i=1}^{n} \xi_i\eta_i \end{aligned} \tag{35}$$

This line goes through the centroid of the points.

Estimation of errors

A statistical estimate of the scatter of the points about the fitted line can be obtained by calculating the mean square deviation

$$\sigma^2_{\text{scatt}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - a - bx_i)^2 \tag{36}$$

where the sum is evaluated with a and b given their 'best' values obtained from the least squares fit. If $\sigma(a)$ and $\sigma(b)$ denote the standard errors on the best values of the coefficients a and b then it can be shown that

$$\sigma^2(b) = \frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{n(n-2)[\xi^2]} = \frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{(n-2)[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2]} \tag{37}$$

and

$$\sigma^2(a) = \bar{x}^2\frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{n(n-2)[\xi^2]} = \bar{x}^2\frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{(n-2)[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2]} \tag{38}$$

$$\text{where } \bar{x^2} \text{ is defined by} \qquad \bar{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \tag{39}$$

These estimates of the errors in the coefficients are generated automatically by the straight line fitting program and should be used to assess the precision of experimental results.

6.2 Weighted straight line fit

In general the experimental points (x_i, y_i) do not have equal precision, and the points must be weighted appropriately. For each value of y_i provide an estimate of the associated standard deviation $\sigma_i(y_i)$. The computer will construct the appropriate weighting factors

$$w_i = \frac{[\sigma_i(y_i)]^{-2}}{\sum_{k=1}^{n}[\sigma_k(y_k)]^{-2}} \tag{40}$$

Best fit parameters

The best fit is again taken to be that which minimises the weighted sum S of the squares of the deviations of the observed points from the straight line

$$S = \sum_{i=1}^{n} w_i(y_i - a - bx_i)^2 \tag{41}$$

subject to the same constraints as for the unweighted fit (equation 33). The following results can then be derived:

$$\begin{aligned} \text{gradient} \qquad & b = \frac{[\xi\eta]}{[\xi^2]} \\ \text{y-intercept} \qquad & a = \bar{y} - b\bar{x} \\ \text{x-intercept} \qquad & x_{\text{int}} = -\frac{a}{b} \end{aligned} \tag{42}$$

where

$$\begin{aligned} \bar{x} &= \sum_{i=1}^{n} w_i x_i & \bar{x} &= \sum_{i=1}^{n} w_i x_i \\ \xi_i &= x_i - \bar{x} & \eta_i &= y_i - \bar{y} \\ [\xi^2] &= \sum_{i=1}^{n} w_i \xi_i^2 & [\xi\eta] &= \sum_{i=1}^{n} w_i \xi_i \eta_i \end{aligned} \tag{43}$$

Estimation of errors

The standard errors on the best values of the coefficients a and b are given by

$$\sigma^2(b) = \frac{\sum_{i=1}^{n} w_i (y_i - a - bx_i)^2}{(n-2)[\xi^2]} = \frac{\sum_{i=1}^{n} w_i (y_i - a - bx_i)^2}{(n-2)[\sum_{i=1}^{n} w_i x_i^2 - \bar{x}^2]} \tag{44}$$

and

$$\sigma^2(a) = \bar{x}^2 \frac{\sum_{i=1}^{n} w_i (y_i - a - bx_i)^2}{(n-2)[\xi^2]} = \bar{x}^2 \frac{\sum_{i=1}^{n} w_i (y_i - a - bx_i)^2}{(n-2)[\sum_{i=1}^{n} w_i x_i^2 - \bar{x}^2]} \tag{45}$$

These estimates of the errors in the coefficients are generated automatically by the straight line fitting routine and should be used to assess the precision of experimental results deduced from the least squares fit. Note that the errors on the coefficients a and b are correlated as may be seen from inspection of equation 41 and equation 42 or equation44 and equation 45.

7 How to write results with errors

Suppose you have taken a set of readings of the length l of a rod. According to the computer the mean is 94.235 814 2 mm, and the standard error in the mean is 0.143 897 6 mm. You do **NOT** write $l = (94.235\,814\,2 \pm 0.143\,897\,6)$ mm. This is because most of the figures in these results are meaningless. If the error is 0.1 mm, there is no justification for quoting the final result to better than 0.01 mm. And the error itself is rarely determined to better than about 10%, so should normally not be quoted to more than two significant figures. Thus, the result should be given as: $l = (94.2 \pm 0.1)$ mm.

Another way of writing this is $94.2(1)$ mm. The convention here is that the number given in brackets is the uncertainty in the last figure quoted in the result itself. This can be convenient if there is a large number of decimal places in the quantity being quoted. For example, the Rydberg constant, which is the quantity determining the energy scale of atomic phenomena, is

$$R = 109\,737.315\,73(3)\ \text{cm}^{-1}$$

It would be cumbersome, though not actually wrong, to write this as

$$R = (109\,737.315\,73 \pm 0.000\,03)\ \text{cm}^{-1}$$

Bibliography

[1] L. Lyons, *A practical guide to data analysis for physical science students*, Cambridge, 1991.

[2] P. R. Bevington, D. K. Robinson, *Data reduction and error analysis for the physical sciences*, 3rd edition, McGraw-Hill, 2002.

[3] L. G. Parratt, *Probability and experimental errors in science*, Dover, 1972.

[4] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd edition, University Science, 1996.

