

# Introduction to Experimental Error

Susan Cartwright

## 1. What is experimental error?

The number that we quote as ‘experimental error’ might be more accurately described as ‘experimental precision’. It is an estimate of the inherent uncertainty associated with our experimental procedure, and is not dependent on any presumed ‘right answer’.

**Example:** Suppose we are asked to measure the length of a block of glass. Our experimental error depends on the method of measurement.

<i>Method</i>	<i>Typical error</i>
cheap ruler	0.5 mm
draughtsman’s ruler	0.2 mm
calipers with vernier	0.05 mm
travelling microscope	0.005 mm
interferometer	0.00001 mm (if $n$ known accurately)

Which of these methods we should actually use depends on how accurately we *need* to know the length of the block: is it just used as a support, or is it part of a precision optical device?

Since we can never measure anything absolutely accurately, it follows that

**any measured quantity has an associated error.**

The associated error is an important part of the measurement, since it allows an outsider to assess the significance of the quoted value in case of any discrepancy with earlier measurements or theoretical predictions.

**Example:** We measure two glass blocks to be 19.0 and 19.5 mm long. Are they really different?

Well—if we used a cheap ruler ( $19.0 \pm 0.5$  mm and  $19.5 \pm 0.5$  mm), we have no reason to believe so. But if we used a travelling microscope ( $19.000 \pm 0.005$  mm and  $19.500 \pm 0.005$  mm), they certainly are!

## 2. Why is experimental error important?

The scientific approach to understanding the world is built on a number of fundamental assumptions and techniques. Two of the most important are:

- *experiments are reproducible* (if you say that experiment X produces result Y, I should be able to do experiment X myself and expect to get result Y);
- *theories are tested by experiment* (if I assert that theory A is an improvement on theory B, I should be able to point out experimental results that are explained by A and not by B, and also correctly predict the results of new experiments that have not been done yet).

Both of these principles only work if we take account of experimental error. If my result Y is 14.09, whereas yours was 14.32, we do not have a problem if the experimental error is  $\pm 0.22$ , but we do if the error is  $\pm 0.02$ ; likewise if I predict that the result of an experiment should be 42, and it turns out to be 41.33, I am happy if the error is  $\pm 0.70$ , but miserable if it is  $\pm 0.07$ .

Because error estimates are fundamental in comparing results, it is crucial that they are properly done: it would be most unfortunate if we rejected a correct theory because of an apparently incompatible experiment, and almost equally unsatisfactory if we clung to a misguided theory because we interpreted a real discrepancy as simply experimental error. (The latter case does tend to correct itself over time, but a number of people have missed out on Nobel Prize-winning new phenomena because they did not take their experimental results seriously enough!)

An awareness of the principles of experimental error is also useful in everyday life: it allows you to make a critical assessment of numerical claims made by politicians, journalists, etc. The principle that any numerical result has an associated error is definitely not restricted to the scientific laboratory.

### 3. Statistical and systematic errors

There are two fundamentally different types of experimental error. *Statistical errors* are random in nature: repeated measurements will differ from each other and from the true value by amounts which are not individually predictable, although the average behaviour over many repetitions can be predicted.

**Example:** Suppose we have a box containing 800 black marbles and 200 white ones, and we attempt to deduce its contents by taking a random sample of ten marbles. An individual sample could easily be 7 black and 3 white, or 9:1, or 10:0, or even 0:10, and it is not possible to predict this. However, it *is* possible to say that the average of many samples will be 8 black and 2 white (assuming that the samples are replaced each time so as not to distort the numbers for subsequent sampling).

Scale reading errors like our previous example belong to this class: if we get 50 people to measure our glass block, we expect to get a range of (slightly) different values. Intrinsically random processes like radioactive decay also belong in this category.

There are standard methods of handling statistical errors. Since they are random in size and direction, they tend to average out: the mean of our 50 glass block measurements will be a more reliable estimate of the length of the block than any of the individual values; in the above example, sampling the contents of the box many times will yield an average of 8 black and 2 white marbles. There is an extensive mathematical literature dealing with statistical errors, and most of the rest of this note will be concerned with them.

*Systematic errors* arise from problems in the design of the experiment. They are *not* random, and affect all measurements in some well-defined way.

**Example:** Unbeknownst to us, our travelling microscope was dropped on the floor during the recent building works in the Hicks Building. This knocked it out of alignment, and all its readings are 1% short. Consequently, the glass block we measured to be  $19.000 \pm 0.005$  mm long is actually  $19.192 \pm 0.005$  mm: a difference far larger than we would expect from the apparent precision of the estimate.

Systematic errors are nasty things. Once recognised, they can often be corrected for—a recalibration of our microscope would disclose the problem and allow us to correct our previous measurements without having to repeat them—but they can be *very* hard to detect. Repeating measurements usually won't help. Some common symptoms are:

- curved lines on a graph where straight lines were expected (especially on log plots);
- nonzero values where zero expected (e.g.  $I$  vs  $V$  plot yielding nonzero current at zero volts);

- inability to reproduce results, even on the same equipment (this could indicate dependence on ambient temperature or pressure, or on how long the apparatus has been switched on).

One should also beware of changing scales on meters and of relying on readings near the upper or lower end of the range (unfortunately these two caveats, while both true, are often mutually incompatible!).

In experiments in the undergraduate lab, systematic errors are often discovered by hindsight (“I didn’t get the right answer—why not?”). Whilst better than nothing, this is not ideal: in the real world you don’t *know* the right answer (if you did, why would you be doing the experiment?). You should therefore try to think about, and list in your lab diary, possible sources of systematic error *before* actually doing the experiment or analysing the data. This may even allow you to eliminate some possible errors at source, e.g. by choosing a suitable meter scale, blocking out stray light, etc. It is also important to *test* any explanation of systematic errors, if possible: take readings with a different meter or a different scale; repeat the analysis using only readings for which the suspected systematic error should be unimportant (if you believe your meter has a 0.01 V zero offset, this is much more serious at 0.05 V than it is at 0.75 V); deliberately heat up an experiment if you suspect that it is sensitive to ambient temperature.

The detection and elimination of systematic errors is something of a black art. You should not feel too disheartened if you miss something that seems obvious once a demonstrator has pointed it out, or indeed if you have a strange result that neither you nor the demonstrator can immediately explain. Your skills in this area will grow as you become more experienced.

#### 4. Statistical errors of measured quantities

We have seen that every experimental datum must have an associated error. In practice, our task is to provide a quantitative estimate of the value of that error. This task divides into two parts: firstly, estimating the errors on directly measured quantities; secondly, using these to calculate the resulting errors on derived quantities.

**Example:** We calculate the current in a particular circuit by measuring the voltage across a standard resistor. We must first estimate the error on the voltage, and then use this to calculate the error on the current. Notice that the uncertainty in the value of the standard resistor introduces a further error in the current: this is a systematic error, because it affects all our current values in a predictable, non-random way.

The method of estimating the value of the error on a particular measured quantity depends on whether we are dealing with a single measurement or with a measurement that has been repeated.

For single measurements, the only guide we have is our knowledge of the experimental setup and our own capabilities. This is unlikely to lead to a particularly accurate estimate of the error, because we may not have sufficiently accurate knowledge of all the relevant factors (for example, a digital display can be read to a precision of  $\pm\frac{1}{2}$  of its last digit, but this does *not* necessarily imply that the instrument concerned is actually capable of that accuracy in the context in which you are using it; likewise, if something is being hand-timed, the dominant source of error may well be the experimenter’s reaction time—a very difficult thing to estimate accurately, as it has both systematic and random components—rather than the accuracy with which the stopwatch can be read). However, even if the final estimate is uncertain by a factor of 2, it is still a reasonable guide to the precision of the final result.

Repeated measurements provide a rather more satisfactory estimate, because we can use the spread in the observed values to derive the characteristics of the underlying random error. It is often convenient to assume that the distribution of measurements around the true value is given by the *normal* or *Gaussian* distribution.

$$P(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right) dx,$$

where  $P(x) dx$  is the probability of obtaining a value between  $x$  and  $x + dx$ ,  $x_0$  is the true value, and  $\sigma = \sqrt{\langle (x-x_0)^2 \rangle}$  is the *standard deviation* or root-mean-square error of  $x$ . The ‘error’ of a single observation is normally quoted as  $\pm\sigma$ .

There are several points to note about this distribution:

- For certain types of error, it can be *proved* that this is the appropriate formula, and it is in fact the underlying distribution assumed in deriving many of the standard results on errors. However, one should be aware that in some cases it is *not* correct, though it is often applied anyway for simplicity. Examples of non-Gaussian errors include digital displays where the intrinsic error is smaller than the least significant digit in the display (in this case we have the ‘top hat’ function, where the probability distribution looks like  $\square$ ), meter readings which involve interpolation between scale divisions (it is known that under these circumstances people do not produce all possible decimals 0.1, 0.2, ..., 0.9 with equal probability), and radioactive decay rates where only one or two counts occur in the specified time interval.
- The estimated error,  $\pm\sigma$ , is *not* the ‘outer limits’ of the variation of  $x$ : in fact, there is a probability of about one third that the true value will lie *outside* the range  $x \pm \sigma$ . This highlights the fact that what we are trying to provide is an estimate of the *probable* error of our experiment, not the largest conceivable error.
- In general, even if we know that the normal distribution is appropriate, we do not know either  $x_0$  or  $\sigma$ . However, these can be reliably estimated from the data if we repeat our measurement a sufficient number of times, as discussed below.

These facts lead to some guidelines for estimating errors:

- Where possible, repeat measurements. If we measure a quantity  $N$  times, then we can estimate  $x_0$  and  $\sigma$ :

$$x_0 = \frac{1}{N} \sum_i x_i = \bar{x};$$

$$\sigma^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2.$$

The quantity  $\sigma$  is then the best estimate of the error on an *individual* measurement  $x_i$ . However, our best estimate of the *true* value is now the mean,  $\bar{x}$ , not any individual measurement. The error on this is the *standard error of the mean*

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} = \left[ \frac{1}{N(N-1)} \sum_i (x_i - \bar{x})^2 \right]^{1/2}.$$

(For some reason, the standard error of the mean does not have a standard symbol, nor is it found as a built-in function on calculators. Fortunately it is simple to calculate if you know  $\sigma$  and  $N$ .)

- If it is not possible to repeat measurements, the best estimate of the true value  $x_0$  is obviously the measured value  $x$ . There is *no information* on  $\sigma$ , which must therefore be estimated from other data, i.e. our knowledge of the experimental setup. In this case it should be remembered that we are trying to estimate the *probable* error, not the worst-case scenario (see above note on the normal distribution).
- A useful special case is that of counting events, e.g. Geiger counts from a radioactive source, photons hitting the CCD of a digital camera, number of events in one bin of a histogram, etc. In this case the underlying distribution is the *Poisson distribution*, which for reasonably large  $N$  (say  $>5$ ) approximates to a Gaussian with  $\sigma = \sqrt{N}$ , i.e. an observation of  $N$  events should be assigned an error of  $\pm\sqrt{N}$ . This error is sometimes called ‘shot noise’.

## 5. Errors on derived quantities

In most experiments, the desired quantity is not measured directly, but is calculated from other quantities which are measured. In this case we need to know how to deduce the error on the calculated result from the estimated errors on the measured values.

### 5.1 Error on $f(x)$ due to error in $x$

Suppose that we have measured  $x$  to be  $X \pm \Delta X$ . The quantity we want to determine is some function  $f(x)$ , and  $f(X) = F$ . What is the error on  $F$ ,  $\Delta F$ , corresponding to  $\Delta X$ ?

Common sense does work: we can evaluate  $f(X+\Delta X)$  and  $f(X-\Delta X)$  to get  $F \pm \Delta F$  (depending on the nature of the function,  $F+\Delta F$  may correspond to either  $f(X+\Delta X)$  or  $f(X-\Delta X)$ ). However, it is often convenient to use the fact that

$$\frac{df}{dx}(x = X) = \lim_{\Delta X \rightarrow 0} \frac{\Delta F}{\Delta X}.$$

If  $\Delta X$  is small, this gives 
$$\Delta F \cong \left| \frac{df}{dx} \right| \Delta X,$$

where the derivative is evaluated at  $x = X$ , and the magnitude is taken because we always take both  $\Delta F$  and  $\Delta X$  to be positive.

**Example:** The following standard formulae are derived using this method:

$$\begin{aligned} \Delta(x^2) &= 2x\Delta x & \text{or} & & \frac{\Delta(x^2)}{x^2} &= 2 \frac{\Delta x}{x} \\ \Delta(x^n) &= nx^{n-1}\Delta x & \text{or} & & \frac{\Delta(x^n)}{x^n} &= n \frac{\Delta x}{x} \\ \Delta(\sin x) &= \cos x \Delta x \\ \Delta(\ln x) &= \frac{1}{x} \Delta x \end{aligned}$$

Note that the angles must be measured in radians, i.e.  $\pm 1^\circ = \pm 0.017$ , and that the formula for the error on a log applies to natural logarithms only (the derivative of  $\log_{10} x$  is **not**  $1/x$ ).

### 5.2 Functions of more than one variable

Suppose we measure two quantities  $x$  and  $y$ , and the value that we need is a function of both,  $f(x, y)$ . How should we combine the errors on  $x$  and  $y$  to get the error on  $f$ ?

We start by assuming that  $x$  and  $y$  are *independent*: that is, the value of  $X - x_0$  (where  $X$  is our measured value and  $x_0$  is the true value) does not in any way influence  $Y - y_0$ . If our measured  $X$  makes  $F$  larger than the true value, our measured  $Y$  may make  $F$  larger still, or it may make  $F$  smaller: we don't know.

Now the square of the standard deviation (the *variance*) is defined as

$$\sigma^2 = \langle (x - x_0)^2 \rangle = \langle x^2 - 2xx_0 + x_0^2 \rangle.$$

If we assume that  $\bar{x} \cong x_0$  (i.e. the mean is a good estimate of the true value), then this simplifies to

$$\sigma^2 = \overline{x^2} - \bar{x}^2.$$

Take the simplest possible function of two independent variables,  $f(x,y) = ax + by$ , where  $x$  and  $y$  are constant. Then

$$\begin{aligned} \sigma_{x+y}^2 &= \langle (ax + by)^2 \rangle - \overline{(ax + by)}^2 \\ &= \langle a^2x^2 + 2abxy + b^2y^2 \rangle - a^2\bar{x}^2 - 2ab\bar{x}\bar{y} - b^2\bar{y}^2 \\ &= a^2(\overline{x^2} - \bar{x}^2) + b^2(\overline{y^2} - \bar{y}^2) + 2ab(\overline{xy} - \bar{x}\bar{y}). \end{aligned}$$

If  $x$  and  $y$  are independent, then  $\overline{xy} = \bar{x}\bar{y}$ , and we have

$$(\Delta f)^2 = (a\Delta x)^2 + (b\Delta y)^2$$

(the variance of the sum is the sum of the variances). The procedure of taking the square root of the sum of squares is called *adding in quadrature*.

What if we have a more complicated function of  $x$  and  $y$ ? If the errors are sufficiently small, we can write

$$f(X, Y) - f(x_0, y_0) \cong \frac{\partial f}{\partial x}(X - x_0) + \frac{\partial f}{\partial y}(Y - y_0)$$

where the partial derivatives are evaluated at our measured values  $X$  and  $Y$ . Since a derivative evaluated at a particular point is just a number (e.g.  $d/dx$  of  $x^2$  evaluated at  $x = 2$  is simply 4), this has, give or take a few constants, the same form as the equation we started with, and we can write

$$(\Delta f)^2 = \left( \frac{\partial f}{\partial x} \Delta x \right)^2 + \left( \frac{\partial f}{\partial y} \Delta y \right)^2.$$

**Example:** The following standard formulae for combining errors are derived from this equation (which looks much nastier than it is!):

$f =$	Apply equation	Simplify
$x + y$	$(\Delta f)^2 = (\Delta x)^2 + (\Delta y)^2$	$(\Delta f)^2 = (\Delta x)^2 + (\Delta y)^2$
$xy$	$(\Delta f)^2 = y^2(\Delta x)^2 + x^2(\Delta y)^2$	$\left(\frac{\Delta f}{f}\right)^2 = \left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2$
$x/y$	$(\Delta f)^2 = \frac{(\Delta x)^2}{y^2} + \frac{x^2}{y^4}(\Delta y)^2$	$\left(\frac{\Delta f}{f}\right)^2 = \left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2$

## 6 Combining data: fits and averages

We have already seen how repeated measurements of a quantity can be averaged to obtain an improved estimate of its true value. This is a simple case of combining data. We shall often meet more complicated cases: for example, our data may consist of single measurements of related quantities, rather than repeated measurements of the same quantity (e.g. values of the current through a circuit for different supply voltages). Alternatively, the measurements may be of the same quantity but with differing precision: for example, if we have two measurements  $1.5 \pm 0.2$  and  $1.70 \pm 0.05$ , it is intuitively obvious that the true value is likely to be between 1.5 and 1.70 but closer to the latter, i.e. the more precise value should be given greater weight in the combination. A number of techniques for combining data to extract the best possible result exist, some of them very complicated. We shall consider only two simple cases.

### 6.1 The weighted mean

Suppose you have measured the same quantity using several different techniques, which naturally give different errors. How do you combine your results to get the best possible estimate of the true value?

The answer is that you need to *weight* your values so that the more precise results have more influence on the final answer. For normally distributed variables with independent errors, the appropriate weight is  $1/\sigma^2$  (i.e. the reciprocal of the variance). This is clear if we consider two values, one of which is the mean of  $N$  measurements each with error  $\sigma$ , and the other a mean of  $M$  measurements with error  $\sigma$ .

- The first measurement is  $\bar{x}_N = \frac{1}{N} \sum_{i=1}^N x_i$ , and its error is  $\frac{\sigma}{\sqrt{N}}$ .
- The second measurement is  $\bar{x}_M = \frac{1}{M} \sum_{i=N+1}^{N+M} x_i$ , and its error is  $\frac{\sigma}{\sqrt{M}}$ .
- Combining both sets of measurements would give  $\bar{x}_{N+M} = \frac{1}{N+M} \sum_{i=1}^{N+M} x_i$ ,  
with error  $\frac{\sigma}{\sqrt{N+M}}$ .

If we define the **weighted mean** as 
$$\bar{x} = \frac{\sum_i x_i / \sigma_i^2}{\sum_i 1 / \sigma_i^2},$$

and its error  $\bar{\sigma}$  by 
$$\frac{1}{\bar{\sigma}^2} = \sum_i \frac{1}{\sigma_i^2},$$

then the weighted mean of the first and second measurement is

$$\frac{\left(\frac{1}{N} \sum_{i=1}^N x_i\right) \cdot \frac{N}{\sigma^2} + \left(\frac{1}{M} \sum_{i=N+1}^{N+M} x_i\right) \cdot \frac{M}{\sigma^2}}{\frac{N}{\sigma^2} + \frac{M}{\sigma^2}} = \frac{\sum_{i=1}^{N+M} x_i}{N+M} = \bar{x}_{N+M}$$

and its error is 
$$1/\left(\frac{N}{\sigma^2} + \frac{M}{\sigma^2}\right)^{1/2} = \sqrt{\frac{\sigma^2}{N+M}} = \frac{\sigma}{\sqrt{N+M}}.$$

So weighting by  $1/\sigma^2$  gives the correct answer in this case. Therefore, since “the mean of  $N$  measurements” is simply a normally distributed variable with a specified standard deviation, it follows that this weighting should give the correct answer for any set of normally distributed variables, provided that the errors are independent.

For our original example of  $1.5 \pm 0.2$  and  $1.70 \pm 0.05$ , the weighted mean comes out to be

$$\frac{1.5/0.2^2 + 1.70/0.05^2}{1/0.2^2 + 1/0.05^2} = \frac{(25 \times 1.5) + (400 \times 1.70)}{25 + 400} = 1.69$$

and its error is  $1/\sqrt{425} = 0.049$ , giving a final result of  $1.69 \pm 0.05$ . Note that the less precise observation contributes almost nothing to the final result.

An alternative expression for the error on the weighted mean is

$$\bar{\sigma}^2 = \frac{\sum_i (x_i - \bar{x})^2 / \sigma_i^2}{(N-1) \sum_i 1/\sigma_i^2},$$

where  $N$  is the number of observations and  $\bar{x}$  is the weighted mean. This estimate uses the scatter of the observations to estimate the error. The two methods should give roughly similar results provided that  $N$  is reasonably large (actually, they are very similar for our example, 0.047 and 0.049, but this is lucky for only two numbers). If they give drastically different values, it is likely that your error estimates are off.

## 6.2 The linear least squares fit

A common situation is the case where two variables (or suitable functions of two variables) are linearly related:

$$y = ax + b,$$

where  $a$  and  $b$  are unknown constants. The data consist of  $N$  pairs of observations  $(x_i, y_i)$ , and the desired physical quantity is  $a$  or  $b$  or both. The problem is to derive the values of  $a$  and  $b$  which best fit the observations, and the corresponding uncertainties  $\Delta a$  and  $\Delta b$ .

The function `linest` will do this for you in Excel (Add Trendline does the same fit, but unhelpfully does not make the errors available). What is `linest` doing, and what assumptions does it make?

The basic principle of `linest` is the *Method of Least Squares*, which involves minimising the quantity

$$\sum_{i=1}^N (y_i - y_i^f)^2,$$

where  $y_i^f = ax_i + b$  is the fitted value of  $y$  corresponding to  $x = x_i$ , and  $y_i$  is the observed value. Now

$$\sum_i (y_i - y_i^f)^2 = \sum_i (y_i - ax_i - b)^2 = \sum_i (y_i^2 + a^2 x_i^2 + b^2 - 2ax_i y_i - 2by_i - 2abx_i).$$

To minimise this we take derivatives with respect to  $a$  and  $b$ :

$$\frac{\partial}{\partial a} \left( \sum_i (y_i - y_i^f)^2 \right) = 2aS_{x^2} - 2S_{xy} + 2bS_x$$

$$\frac{\partial}{\partial b} \left( \sum_i (y_i - y_i^f)^2 \right) = 2bN - 2S_y + 2aS_x$$

(where we have used the notation  $S_x = \sum_i x_i$  etc.), and we require both of these derivatives to be zero:

$$S_{xy} = aS_{x^2} + bS_x;$$

$$S_y = aS_x + bN.$$

Solving this pair of simultaneous equations for  $a$  and  $b$  gives

$$a = \frac{NS_{xy} - S_x S_y}{NS_{x^2} - S_x^2};$$

$$b = \frac{S_{x^2} S_y - S_x S_{xy}}{NS_{x^2} - S_x^2}.$$

Deriving the errors is a bit more tedious, but you end up with

$$\frac{(\Delta a)^2}{N} = \frac{(\Delta b)^2}{S_{x^2}} = \frac{1}{N-2} \frac{\sum_i (y_i - y_i^f)^2}{NS_{x^2} - S_x^2}.$$

This is what Excel does. There are a few things to note about this method of fitting a straight line:

1. We have worked in terms of minimising  $(\Delta y)^2$ . This basically means that we assume that  $y$  has larger errors than  $x$ . If this is not true, it is best to rearrange the equation so as to reverse the roles of  $x$  and  $y$ .
2. **We have assumed that the errors in  $y$  are all about the same size.** If this is not true, we need a *weighted least squares fit*, which is similar to the weighted mean we discussed in the previous section. **Excel does not have a built-in weighted least squares fit**, but it is not difficult to design a spreadsheet to do it (an example is given in Les Kirkup's *Experimental Methods*).
3. We have assumed that the errors in  $y$  are *independent*. If there is a systematic error, this will not be properly taken into account in the fit. It is best to fit using only the statistical errors, and then decide what effect the systematic error will have (to take a simple example, if the measured values of  $y$  have a common calibration error of 5%, then the systematic error on  $a$  will also be 5%, because *all* the values of  $y$  will move up and down together if the calibration is changed).

It is also worth noting that you should *always* look at your data before fitting. It is of course possible to fit the best straight line to a set of data regardless of whether the points in question actually resemble a straight line! In particular, systematic errors and approximations often have more effect at one end of the range, leading to a straight line that levels off at low or high values: in this case you will get a more accurate result by restricting the fit range to the part of the graph that actually looks linear. Similarly, a single very bad measurement can

have a big effect on the final gradient, and in some cases it would be appropriate to omit such an outlier from the fit.

You should also look at your data *after* fitting. If the errors on  $y$  are Gaussian, we expect that about  $2/3$  of the points should lie on the fitted line within their error bars, and only one point in 20 should be more than twice its error bar off the line (no points should be more than three error bars off). If too many points lie off the line, you have underestimated your errors, or there is a systematic effect. If too *few* lie off the line, you have overestimated your errors, or the errors are not really independent. Note that if you have 20 points, all lying on the fitted line within their error bars, there must be something *wrong*: the probability of this happening with accurately estimated Gaussian errors is actually very small (about 1 in 2000).

## 7 Quoting errors

Last but not least, we must consider how best to *present* our results. The aim here is to provide the information—the experimental value and its associated error—in the way which is most easily grasped by a reader who is not necessarily familiar with the details of the experiment.

### 7.1 Quoting the numbers

Error estimates are usually just that—*estimates*—and should not be quoted to unjustified precision. Usually one significant figure is enough, although if that one figure is 1 or 2 most people would add a second:  $14.2 \pm 0.6$ , but  $14.18 \pm 0.17$ . You may want to keep 2 significant figures if you are going to be using this value in subsequent calculations:  $(14.18 \pm 0.56)^{1/2} = 3.77 \pm 0.07$ , whereas  $(14.2 \pm 0.6)^{1/2} = 3.77 \pm 0.08$ .

Having decided how many figures to use in your error, round the main value to the same number of decimal places (*not* the same number of significant figures!): for example, Planck's constant is  $(6.62606876 \pm 0.00000052) \times 10^{-34}$  J s (source: Particle Data Group, Phys. Rev. **D66** (2002) 010001). If using scientific notation, always use the same exponent, as above:  $6.62606876 \times 10^{-34} \pm 5.2 \times 10^{-41}$  J s is numerically identical, but much less easy to interpret. Don't over-round the main value:  $6.626 \times 10^{-34} \pm 5.2 \times 10^{-41}$  J s is *wrong* (in this case the imprecision is dominated by the rounding error,  $\pm 0.0005 \times 10^{-34}$  J s).

In summary:

- quote error to not more than two significant figures;
- quote value so that it has the same number of decimal places as the error;
- when using scientific notation, quote value and error with the same exponent;
- when you report a numerical value, **always include its error** (an experimental result is completely useless without an error estimate).

### 7.2 What to say about errors on primary (measured) quantities

Always work out your errors *as you go along*: don't wait until you have a final answer and then think, "What is the error on this?" The main reason for this is that it will improve the quality of your results: if you are doing the errors as you go, you will be able to see which of the various sources of uncertainty is having the biggest effect, and this gives you a chance to do something about it.

In discussing your error estimates, whether in your lab diary or in a report, you should:

- carefully state the source(s) of uncertainty, how the values quoted were estimated, and whether the resulting error is statistical or systematic;

- if the quoted value is a mean of  $N$  observations, state  $N$ , and make clear whether the quoted error is the standard deviation (i.e. the error on an individual measurement) or the standard error of the mean;
- combine different sources of statistical error (for example photon shot noise and electronic noise in a CCD image) by adding in quadrature, i.e.  $\sigma_{\text{tot}}^2 = \sigma_1^2 + \sigma_2^2$ ;
- quote statistical and systematic errors separately, e.g.  $1.57 \pm 0.09$  (stat.)  $\pm 0.05$  (syst.).

### 7.3 What to say about errors on secondary (derived) quantities

In reporting to the outside world (a lab report or poster), you should *not* need to derive or even quote the basic error formulae such as those in section 5. Your intended readers are trained scientists, who will *expect* you to use these formulae. Similarly, do not waste space writing down your arithmetic: the readers can work this out for themselves. What you do need to do is

- make sure that you have explained how the derived quantity is related to the measured quantities (i.e. give the equation);
- carefully explain any systematic errors, and how you have dealt with them (e.g. by restricting the range over which you have fitted the data).

### 7.4 How to compare experiment and prediction or accepted value

In most lab experiments, the result you obtain can be compared directly with a theoretical expectation or a ‘book value’. In making a comparison, it is *essential* to take the errors into account: that’s what they’re for. The basic idea is to subtract the book value from your value, calculate the error on the difference (including the error on the book value, if there is one), and see how far from zero the result is. Roughly speaking,

- if the difference is zero within one error bar (e.g.  $1.3 \pm 1.4$ ), the results are in *good agreement*;
- if the difference is between one and two error bars from zero (e.g.  $1.3 \pm 0.8$ ), the results are *consistent* (there is about one chance in three that this could happen by chance);
- if the difference is between two and three error bars (e.g.  $1.3 \pm 0.5$ ), you probably have a problem (the chance of this is about one in 20);
- if the difference is bigger than three error bars (e.g.  $1.3 \pm 0.3$ ), you *definitely* have a problem (for normally distributed errors, the chance of this is less than 1%).

If you have a problem, *do not ignore it*. It is simply not acceptable to say “the difference between my result of  $1094.2 \pm 0.3$  and the book value of 1095.5 is  $1.3 \pm 0.3$ , but this is less than 0.1%, so my value is very close.” Your value is 4.3 error bars off, which is so wildly improbable that it will be off the end of most statistical tables. Either your error estimate is off by at least a factor of 2, or you have a genuine problem, perhaps a systematic error.

It is also not acceptable to say “the discrepancy is due to inaccurate measurements.” The inaccuracy of your measurements should already be included in your error estimate: if it isn’t, you haven’t done your error estimating properly. Acceptable explanations for a discrepancy might include

- different conditions (“the book value is quoted at  $0^\circ$  C, whereas my measurements were carried out at room temperature”);

- dubious approximations (“the theoretical prediction uses the small angle approximation, but my measurements were carried out at an angle of  $10^\circ$  (0.175 radians), which is not that small”);
- possible systematic effects (“it was not possible to calibrate the meter accurately, because we did not have access to a standard source”).

More quantitative estimates of probability (*confidence limits*) can be made by using statistical tables which give the integral of the normal distribution. There are some online tools that do this, for example Quickcalcs (<http://graphpad.com/quickcalcs/probability1.cfm>) and StaTable (<http://www.cytel.com/statable/>). Unfortunately, both of them regard a discrepancy of 4.3 standard deviations as completely beyond the pale: Quickcalcs doesn't go out that far, and StaTable informs me that the probability of  $1.3 \pm 0.3$  is 0.0 (the best it will give me is  $1.3 \pm 0.33$ , which has a probability of 0.008%).

When quoting book values, always specify the book: some accepted values do evolve over time as measurements improve (for example, the value of the Hubble constant is considerably better known now than it was five years ago), some depend on the exact convention used in the definition, and there is always the possibility that your book has a misprint at the critical point! Therefore, do not say “the book value for the specific heat of iron is  $460 \text{ J kg}^{-1} \text{ K}^{-1}$ ,” but rather “the accepted value of the specific heat of ductile iron at  $23^\circ\text{C}$  is  $460 \text{ J kg}^{-1} \text{ K}^{-1}$  (source: [http://metalcasting.auburn.edu/data/Ductile\\_Cast\\_Iron/Ductile.html](http://metalcasting.auburn.edu/data/Ductile_Cast_Iron/Ductile.html))”. The second version gives the source, and also provides more specific information (the value is noticeably temperature sensitive, and ductile iron is probably different from, say, wrought iron).

## 8 Further reading

1. Les Kirkup, *Experimental Methods* (John Wiley, 1994): recommended in the first year lab, a useful introductory textbook that covers keeping a lab diary, writing a lab report, and spreadsheet recipes for common calculations (such as least squares fits) as well as basic statistics and error calculation.
2. Roger Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (John Wiley, 1989): a somewhat more advanced, but still easy to read, textbook written by a Manchester University particle physicist.
3. Louis Lyons, *A Practical Guide to Data Analysis for Physical Science Students* (Cambridge University Press, 1991): similar in level to Kirkup, but more focused on data analysis; written by another particle physicist (well, we use statistics rather more than some other branches of experimental physics!).