The Analysis of Errors -- a practical guide

I. Introduction

Physicists are concerned not only with measurements, but also with the evaluation of the precision of the measurement. The questions inevitable arise: “How well has a theory been verified? How well can I believe a prediction? Can I use that data with confidence?” Hence, the importance of the subject of error analysis is.

The term *error* in its simplest application describes the deviation of a measurement from the correct value. Very often, however, the correct value is not known. The term error is then often used in the sense of an estimated error, or precision index, which the experimenter assigns as carefully as he can to his measurement. Another term, *discrepancy*, is often useful to describe the deviation of one measurement of the same quantity from another, or of the measurement from the prediction of a theory. The use of the term "discrepancy" in such cases is less confusing than the term "error".

The determination of most physical qualities can be represented schematically in this way:

INPUT DATA	⇒	CALCULATIONS OR GRAPHS	⇒	RESULTS
(Error estimates)		(Propagation of errors)		(Estimation of overall error in result)

As the diagram brings out, there are two basic parts to the analysis of error: the estimation of the accuracy of the input data and the propagation of these errors into the result to give an overall error estimate. In what follows, two main types of errors are discussed, together with some suggestions for evaluating and reducing them, and then some rules are given for the propagation of errors.

II. Sources of error and their significance

1) Random Errors

Random errors occur where repeated measurements of a quantity give different results, usually with as many positive as negative fluctuations. If a large number of measurements are tabulated according to the number which fall in a given measurement interval, one obtains a histogram which represents the distribution of errors. By far the most common error distribution is the Gaussian distribution illustrated in Figure 1.

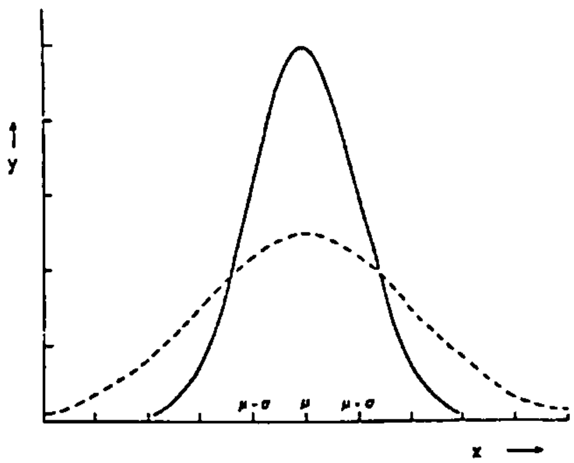

Figure 1 The solid curve is the Gaussian distribution. The distribution peaks at the mean μ, and its width is characterized by the parameter σ. The dashed curve is another Gaussian distribution with the same value of μ, but with σ twice as large as the solid curve.

An almost universally used measure of the spread of an error distribution is the standard deviation defined as:

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}} \, . \tag{1}$$

Here x_i's are the individually measured values,

$$\mu = \bar{x} = \frac{\Sigma x_i}{n} \tag{2}$$

is the mean value, and the sum is taken over all n measurements. For a Gaussian distribution, about 31.7% of the measurements fall further from the mean, either way, than the standard deviation. Only 5% fall outside $\pm 2\sigma$.

However, there are theoretical arguments for replacing n in Eq. (1) by n - 1 and defining the standard deviation σ as:

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} \tag{3}$$

Equation (3) is a better definition theoretically, especially for the case where n is small. Obviously, when $n \to \infty$, both definitions are equal.

The difference between the two definitions shown in Eqs. (1) and (3) is almost always numerically not too significant, thus for most purposes we can use either definition to calculate the σ. However, it may be better always to use the more conservative (i.e. larger) definition Eq. (3) due to the fact that it is a more mathematically rigorous definition; in any case, your laboratory report should state clearly which definition is being used, so that the TAs can check the calculations.

In actual experimental situations, neither the mean nor the standard deviation is known. In principle, one could repeat the measurement a very large number of times and then the mean would approach the correct value. In practice, it is hardly ever worth taking more than 10 measurements because the accuracy achieved by taking the mean of a sample of n measurements improves only as $1/\sqrt{n}$. Greater accuracy is usually sought by improving the method of measurement.

Repetition of a given measurement is nevertheless extremely valuable because it gives an estimate of the precision. Often it suffices just to see how the result varies from one measurement to the next; errors are in any case an approximate business. The standard deviation can be calculated according to the definition given above; if after doing this for the individual measurements the mean of the sample is calculated, the precision of this operation may also be quoted as a standard deviation, in this case approximately equal to the original standard deviation *divided by* $\sqrt{n}$ (for more discussions, see section III).

Sometimes a set of measurements has to be taken under prescribed conditions. The error of one of the quantities involved in producing the prescribed conditions can be estimated by noting to what extent the quantity can be varied without disturbing the prescribed condition.

2) Propagation of Random Errors

The experimental determination of some quantity usually involves the measurement of several observables which are then inserted in an equation to give the desired result. Each observable may have an error associated with it, and these errors sue said to *propagate* into the result.

A general formula for the error in R where R is a function of X_1, X_2, X_3, ……X_n is:

$$\Delta R = \sqrt{\left(\frac{\partial R}{\partial x_1}\Delta x_1\right)^2 + \left(\frac{\partial R}{\partial x_2}\Delta x_2\right)^2 + \ldots\ldots + \left(\frac{\partial R}{\partial x_n}\Delta x_n\right)^2} \qquad (4)$$

The errors from each of the quantities X have been added in quadrature (root mean square) because for uncorrelated errors the signs of the errors are equally likely to be plus or minus. Specific cases are discussed below in Section III.

3) Systematic Errors

Examples of systematic errors are:

(1) Defective or inaccurate measuring instruments
(2) Improper use of instruments
(3) Faulty experimental methods
(4) Neglect of such effects as temperature, humidity, wind resistance, friction, etc.
(5) Bias on the part of the observer.

Very often, systematic errors are avoided by calibrations and corrections. The knack of spotting systematic errors is one of the attributes of a good experimenter, and it must be developed through practice.

Examples of systematic errors are the following. In measuring the length of a rod, the meter stick might not have been properly marked, or the wood may have expanded due to humidity. In measuring the diameter of a cylinder, the cylinder may not be perfectly round. The speed of a falling object may depend on air resistance as well as the acceleration of gravity.

Many times one must depend on the reputation of the manufacturer and the condition of the instrument. The meter stick could be checked against another one, and the experiment could be repeated in a dry room. The cylinder could be measured across several different diameters. The speed of denser or lighter objects could be measured. Usually it helps to vary the conditions of the experiment and to use another observer.

As opposed to random errors, systematic errors tend to be of a definite sign, or more or less definite magnitude. Repeating the experiment without changing the conditions does not help to reduce the systematic errors.

4) Significant Figures

As many figures should be retained in a measurement as are meaningful according to the estimated error. Thus, 2.017676 ± 0.01 has four extra figures which add nothing to the result. It should be written as 2.02 ± 0.01.

The following table shows how ambiguous use of figures can be avoided:

Measured Value	Number of Significant Figures	Remarks
2	1	Implies = 25 percent precision
2.0	2	Implies = 2.5 percent precision
2.00	3	Implies = 0.25 percent precision
0.136	3	Leading zero is not necessary, but it does make the reader notice the decimal point.
2.483	4	
2.483×10^3	4	
310	2 or 3	Ambiguous. The zero may be significant or it may be present only to show the location of the decimal point.
3.10×10^2	3	No ambiguity
3.1×10^2	2	

III. Discussion of Errors in Lab Write-Ups

Most experiments have several parts, with each part involving the measurement of certain input data followed by a calculation of the experimental value of some physical quantity (call it R) as deduced from these data. Often the same quantity will be measured in two different ways (R_1, R_2) or the accepted value will be known (R_A) or the prediction of a theory (R_T) will be available.

When the input data are recorded, the student should write on the data sheet the estimated error. This is usually written like $X \pm \Delta X$. There are no universal rules for estimating errors. The measurement may be repeated, the apparatus may be repeatedly adjusted to the prescribed condition, etc.

In discussing the errors, the basis for the estimation of the errors should be mentioned in the write-up. Sample calculations should also be given to show how the input data errors are combined (the rules for combining errors are given below) to give the estimated error of the result (ΔR). Then the result should be compared with other measurements of the same quantity, with an accepted value, or with a theoretical prediction (R_T). The difference, $R - R_T$, for example, is called discrepancy. The big question is: do these two values agree within estimated errors? They do if $R - R_T$ is of the same order, or not much larger than, the sum of ΔR and ΔR_T. One would then be able to conclude that the theory has been verified to the precision of the present experiment. The foregoing can be summarized by Figure 2:

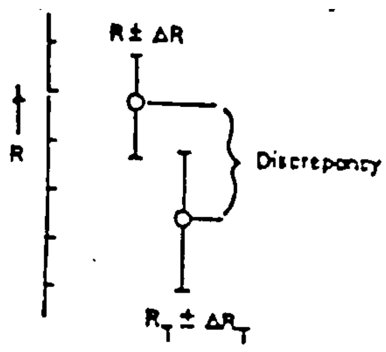

Figure 2 Schematic representation of errors and discrepancy

The agreement (or disagreement) obtained in the above fashion can then be discussed in terms of possible undetected systematic errors.

Reducing Uncertainty

There are a number of experimental procedures available to reducing uncertainties although often only some or even only one will be appropriate to a particular situation. Experience will enable you to select.

(1) Repeat the measurements

It is useful when a series of measurements, rather than a single measurement, are taken. The idea is to take a number of repeated measurements (X_1, X_2 ... X_n, say) and then to find the mean.

$$\overline{X} = \frac{X_1 + X_2 + \dots\dots\dots + X_n}{n}$$

The uncertainty can then be estimated by computing the standard errors on the mean of the random errors: $s = \frac{\sigma}{\sqrt{n}}$, where σ is the standard deviation of the sample:

$$\sigma = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n-1}}\ .$$

There are a number of things to note here:

a) It makes no sense to compute s if n less than 5.
b) Repeating a measurement five times is time consuming and may not be necessary, if the associated uncertainty is small compared with others in the experiment.
c) If this procedure is followed, it becomes unnecessary to consider the individual uncertainties as contributions to s.
d) An alternative estimate of the error is to take 2/3 of the total spread. While this is less precise it is often more sensible.

(2) Graphing and regression analysis

A powerful way of increasing the reliability of any result is to vary the conditions in such a way that the data can be plotted as a graph, in a way that allows the final result to be calculated using the graph (gradient, intercept). This avoids taking the same measurement too many times and also provides considerably more information.

It is still essential in any analysis to obtain an estimate of the associated uncertainties and there are a number of ways of doing this. Two methods are described here: (i) graphical and (ii) statistical.

(i) The first step in the graphical method is to put "error bars" or "error crosses" on the points, as shown in Fig. 3. These are used to represent the individual uncertainties in the measurements or quantities plotted.

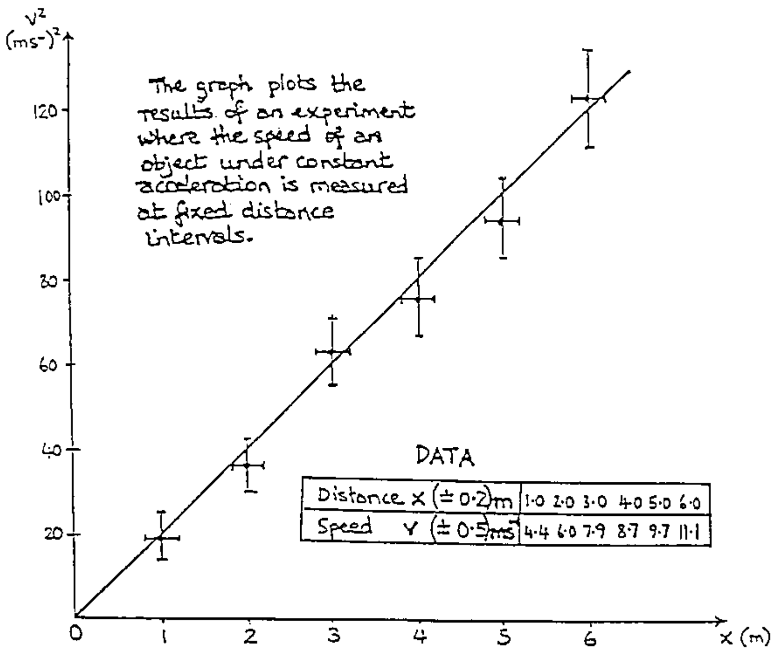

Figure 3 Presentation of individual uncertainties on the measurement points

The uncertainty in the gradient can now be estimated by drawing the line with the steepest gradient that still passes through all the error crosses, and then the line with the flattest gradient that does this. The best fit will usually be the mean of these two, and the gradients of these lines will represent the uncertainty limits. This is shown in Fig. 4 below.

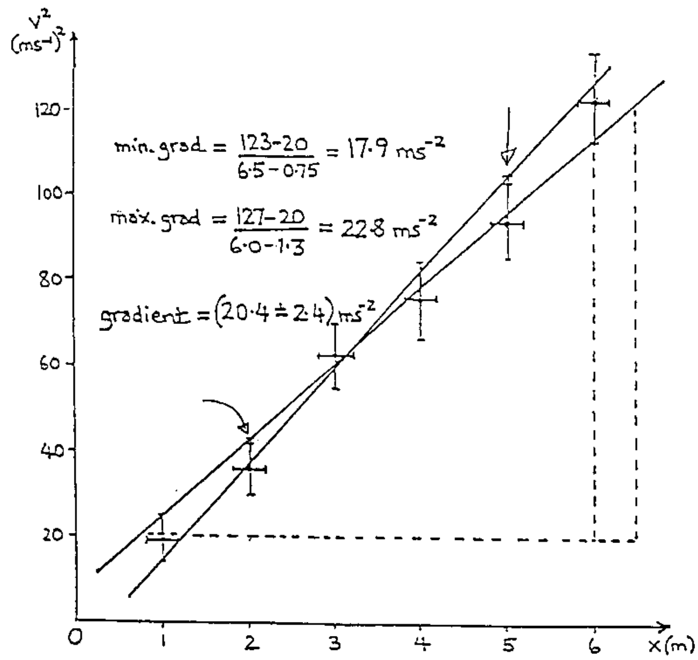

Figure 4 Determination of the best fit slope

(ii) The most common statistical method is the *Least Squares Regression Method*. A computer programme is available for this. For a group of data points (x_1,y_1), (x_2,y_2),...(x_i,y_i),..., we can fit a straight line relationship

y = m x + c

where the slope $m = \dfrac{\Sigma\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\Sigma\left(x_i - \bar{x}\right)^2}$

with uncertainty $\pm \Delta m = \sqrt{\dfrac{\Sigma\left(y_i - mx_i - c\right)^2}{(n-2)\Sigma\left(x_i - \bar{x}\right)^2}}$

and intercept $c = \bar{y} - m\bar{x}$

with uncertainty $\pm \Delta c = \sqrt{\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{\Sigma\left(x_i - \bar{x}\right)^2}\right)\dfrac{\Sigma\left(y_i - mx_i - c\right)^2}{(n-2)}}$

(3) Systematic errors

Given care in avoiding the more common systematic errors, parallax for example, the biggest difficulty involved is detecting whether there are any. Often the only way to do this is to use the equipment to obtain a result for which there already is an accepted value. The test is then whether this value lies inside or outside the experimentally predicted range. If it lies outside, either the random error has been underestimated or there is a systematic error present.

For example, in six identical experiments to measure the gravitational acceleration, the following results were obtained:

g (m/s^2)	10.4	10.5	10.3	10.4	10.6	10.2

Taking the accepted value for g to be 9.8 m/s^2, estimate the random and systematic errors associated with this experiment.

NOTE: be careful to distinguish between accepted values and nominal values.

Accepted values will be found in standard data texts, for example, densities, conductivities etc. Copies of these texts are available in the laboratory.

Nominal values are found stamped on pieces of equipment, such as standard resistors, capacitors, klystrons etc. In general, these values will always have a non-specified tolerance which may well be 10% or more.

1) Combining Errors

(1) Single measurement

When errors (e_1, e_2,...) are combined for repeated, and apparently reproducible measurements of a simple quantity, the resulting uncertainty is:

$$e = \sqrt{\frac{e_1^2 + e_2^2 + \ldots + e_n^2}{n}},$$

provided the individual errors are independent, this gives a more realistic estimate than just summing.

(2) Combined result

In many cases a particular result will be obtained by combining a number of measurements of different quantities in some way.

If independent measurements (X ± ΔX) and (Y ± ΔY) are combined to give a result R, the associated uncertainty ΔR will be given by:

$$\left.\begin{aligned} R &= X + Y \\ R &= X - Y \end{aligned}\right\} \qquad \Delta R = \sqrt{\Delta X^2 + \Delta Y^2}$$

$$\left.\begin{aligned} R &= XY \\ R &= X / Y \end{aligned}\right\} \qquad \frac{\Delta R}{R} = \sqrt{\left(\frac{\Delta X}{X}\right)^2 + \left(\frac{\Delta Y}{Y}\right)^2}$$

$$R = kX \qquad \frac{\Delta R}{R} = \frac{\Delta X}{X} \qquad \text{(k is a constant)}$$

$$R = X^n \qquad \frac{\Delta R}{R} = n\frac{\Delta X}{X}$$

$$R = XY^n \qquad \frac{\Delta R}{R} = \sqrt{\left(\frac{\Delta X}{X}\right)^2 + n^2\left(\frac{\Delta Y}{Y}\right)^2}$$

For more than two quantities the equations are simply extended.

2) Example

During an air track experiment, the experimenter placed an object of unknown height at the foot of the air track so that the air track was made slightly inclined. A glider was placed on the air track so that it fell freely along the track. The motion of the glider was then recorded by the Sonic Ranger:

Speed (m/s)	Time (s)
0.45 ± 0.06	1
0.81 ± 0.06	2
0.91 ± 0.06	3
1.01 ± 0.06	4
1.36 ± 0.06	5
1.56 ± 0.06	6

Using the above guidelines, the data are graphed in Fig. 5.

Using the statistical method of least square fit, the following results are obtained for the best fit straight line y = mx + c:

m = 0.21 ± 0.02
c = 0.29 ± 0.08 * These results are calculated by Excel

By Newtonian mechanics, we know that v = u + at. Thus, by plotting v against t, we obtain the value of a = 0.21 ± 0.02 ms^{-2}.

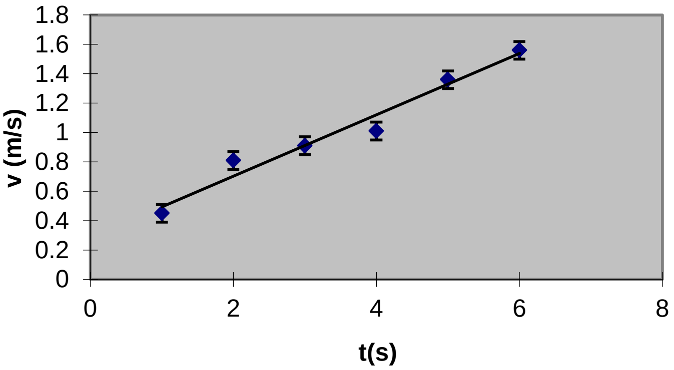

Figure 5 Measured velocity as a function of time t

Provided the height of the object is very small compared with the length of the air track, the height of the object can be approximated by:

$h = a\, l / g$ where h = height of the object
a = acceleration of the glider
g = acceleration due to gravity
l = length of the air track

Given that $l = 1.75 \pm 0.05$ m and that $g = 9.8\ \mathrm{ms}^{-2}$ (an accepted value), the height of the object can be approximated as:

$$h = 0.21 \times 1.75 \div 9.8 = 0.0375 \text{ m}$$

$$\Delta h = h \times \sqrt{(0.02/0.21)^2 + (0.05/1.75)^2} = 0.004 \text{ m } *$$

So, the height of the object is 3.8 ± 0.4 cm.

* Note: as the error of l is very small compared with a, it may be neglected from the error estimation.

IV. References

1. P. R. Bevington, *Data Reduction and Error Analysis in the Physical Sciences* (McGraw Hill, 1968).
2. A. J. Lyon, *Dealing with Data (*Pergamon, 1970).
3. J. Topping, *Errors of Observation and their Treatment* (Inst. Phys. Lond., 1962).
4. H. D. Young, *Statistical Treatment of Experimental Data* (McGraw Hill).
5. J. R. Taylor, *An introduction to error analysis: The study of uncertainties in physical measurements* (University Science Books)