M. Longo

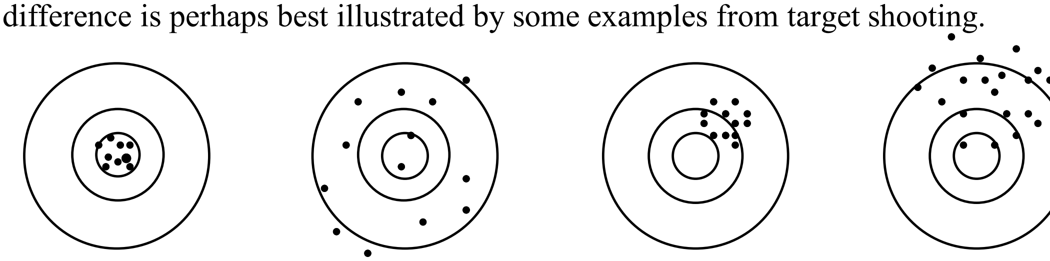
ERRORS AND THE TREATMENT OF DATA

Essentially all experimental quantities have an uncertainty associated with them. The only exceptions are a few "defined" quantities like the wavelength of the orange-red light from Krypton-86, which is defined to be 1/(1,650,763.73) of a meter. In other words, we have defined our unit of length in terms of this wavelength. The uncertainty in physical measurements becomes of crucial importance when comparing experimental results with theory. As you all know, the basis of the "scientific method" is to test our hypotheses against experimental data. Unless you take the experimental errors into account in the physics labs you'll very quickly find that you have disproved all the "laws" of physics! Obviously, we'd rather not let you go away with that impression, so we want you to learn how to estimate how large an uncertainty or error to attach to your results due to uncertainties in your measurements.

The assignment of probable errors to physical data is not easy. Some sources of error can be estimated fairly accurately; others may be difficult or impossible to estimate. The history of physics has many notorious examples of experimenters who have grossly underestimated the errors in their measurements. This is partly a result of human nature–you like to think your experiment is more accurate than anybody else's–and partly a result of lack of knowledge. Sometimes there are sources of errors the experimenters didn't know about; sometimes they knew about it but didn't know how to estimate the effect on their results properly.

Systematic *vs.* Random Errors

Random errors are those produced by unknown and unpredictable variations in the experimental situation. Systematic errors are errors associated with a particular instrument or experimental technique (though not all errors associated with an instrument are systematic). The difference is perhaps best illustrated by some examples from target shooting.



The random errors might be due to variations in the cartridges, jitter while aiming, *etc.* The systematic errors might be wind, misaligned sights, or a consistent bias in aiming. Note that random errors become smaller if the data are averaged over many tries or measurements, while systematic errors do not. In other words, we can decrease random errors by taking many measurements and averaging, but we must combat systematic errors in other ways.

The Normal Distribution, Mean, and Standard Deviation

Suppose we consider a measurement whose result can take on a continuous range of values. To be concrete, let us imagine a very simple experiment. We want to measure the time it takes a ball to fall 1.00 meter. To get an accurate value we use a good stopwatch and repeat the measurement 200 times. Figure 1 shows the results of our hypothetical experiment in the form of a histogram.

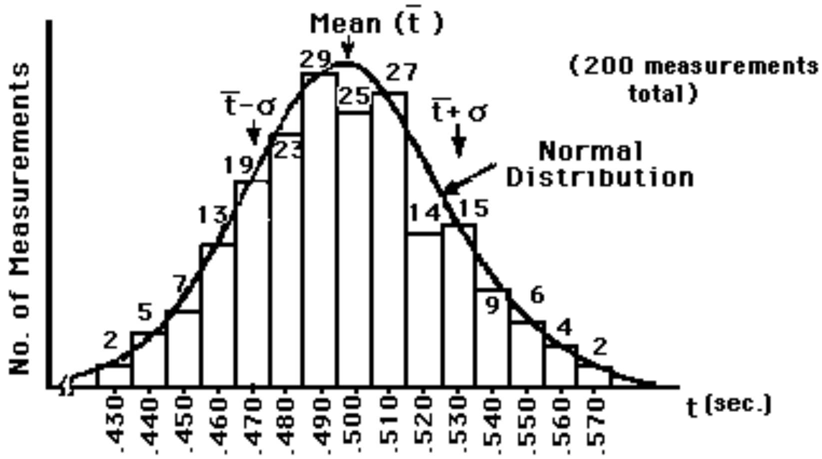


Fig. 1

The vertical height of each rectangle or "bin" gives the number of measurements that lie within the range of the bin. For example, there were 29 measurements with fall times between 0.485 and 0.495 sec.

The distribution in Fig. 1 is somewhat idealized, but is typical of what real data from a well-designed and executed experiment might look like. The most notable features are:

(1) The values are clustered about a well-defined mean value which is close to the most probable value (the value of t where the distribution has its maximum). The arithmetic mean of the t values in Fig. 1 is approx. 0.497 s.

(2) Values that are far from the mean are very unlikely.

(3) The distribution is reasonably symmetric about the mean. There is no obvious skewing toward the high or low side.

If we took many, many measurements and made the bins very fine, our histogram might begin to look like the smooth, bell-shaped curve. This curve is the limiting case in an ideal situation. It is referred to as the normal or Gaussian distribution. Measurement errors that follow this distribution, are said to be normally distributed. The mathematical form of the normal distribution is really not very important because, in a given experiment, you cannot prove that the measurements will follow a normal distribution. Nevertheless, a distribution resembling the normal distribution is usually found, and it is usually assumed that a normal distribution is appropriate.

The bell-shaped curve and, to a good approximation, the histogram of Fig. 1 can be characterized by two quantities, the mean value and the width. The mean $\bar{t}$ of the measured times is just the arithmetic average of the data,

$$\bar{t} = \frac{1}{N}\left[t_1 + t_2 + ... + t_N \right] \equiv \frac{1}{N}\sum_{i=1}^{N} t_i \qquad (1)$$

Here t_1, t_2, . . . are the measured times, the symbol Σ stands for a sum, and N is the number of measurements. (For the normal distribution, the mean is defined in terms of an integral analogous to Eq. 1.) The width of the distribution can be defined in various ways – for example, the full width at half maximum or the root mean square (rms) deviation σ that is defined as

$$\sigma = \left[\frac{1}{N}\sum_{i=1}^{N} (\bar{t} - t_i)^2 \right]^{1/2} \qquad (2)$$

where $\bar{t}$ is the mean from Eq. 1. The rms deviation turns out to be the most common, and we shall accept it as our definition; σ can be thought of as the probable error or uncertainty in one measurement. It is also called the standard deviation of the measurement. We shall generally refer to it as the standard deviation and use the symbol σ. Generally then we will use the mean, defined in Eq. 1, as the best estimate of the measured quantity, and the standard deviation of the mean $\bar{\sigma}$, defined in Eq. 3 below, as the best estimate of the uncertainty of the mean.

If the measurements follow a normal distribution then ideally 68.3% of the measurements lie within $\pm 1\ \sigma$ from the mean. Thus, from Figure 1, which contains 200 measurements, we can estimate σ by counting off 68 measurements in either direction from the mean. This includes a band of width approx. ±0.030, so the standard deviation per measurement is approx. 0.03. Note that we could get a more precise value of σ by numerical calculation from Eq. 2, but the increase in precision of σ is insignificant. In other words, we shouldn't feel obliged to estimate σ to very high accuracy. In practice, it is safest to use the histogram method for estimating σ when possible because it gives a chance to judge whether the data look "normally" distributed. We might be tempted to discard a measurement that lies many standard deviations from the mean. (We shall not discuss the correctness of this procedure; the point is that it is often done.) It is important to realize that σ is a measure of the probable uncertainty of one measurement —*i.e.*, if we make one measurement it has a 68% probability of being within 1σ of the mean value. The uncertainty in the mean is much smaller than σ because we have made many measurements. For N measurements the standard deviation of the mean $\bar{\sigma}$ is

$$\bar{\sigma} = \sigma/\sqrt{N} \tag{3}$$

This assumes that the measurements are independent and uncorrelated. In the example of measuring the fall time of a ball, if we started and stopped two clocks with the same switches the measurements of the two clocks would be strongly correlated; the amount of correlation would depend on how good the clocks were. (The better they are, the stronger the correlation.)

The result of a series of measurements of a quantity A and its error or uncertainty are usually written in the form $\bar{A} \pm \bar{\sigma}$. In the example above, the mean time was 0.497 sec and the standard deviation of the mean would be $0.03/\sqrt{200}$, so we would write

$$\bar{t} = 0.497 \pm 0.002 \text{ sec}$$

Ideally this means that the "true" value of $\bar{t}$ has a 68.3% chance of lying between 0.495 and 0.499 sec. Two results are considered to be consistent with each other if they are within 1 or 2 standard deviations of each other. Obviously some judgment is required.

The above discussion assumes that all of the measurements in an experiment are of equal intrinsic accuracy. If some measurements are better than others, the better ones should have a higher weight in computing the mean. The calculation of the weighted mean and probable errors in this kind of a situation is discussed in many references.

Experiments Whose Outcome Is an Integer: The Square Root Rule

Often the result of an experiment or measurement is an integer—for example, the number of mice out of an initial sample of 100 that die within one year or the number of radioactive nuclei out of a sample that decay in one second. The standard deviation of the number of such "events" (deaths, decays, or whatever) can be estimated by the square root rule. If N is the number of events, the standard deviation in N is

$$\sigma = \sqrt{N} \tag{4}$$

For this to be an accurate estimate, the following conditions must be satisfied. (The better they are satisfied, the better the estimate of σ.)

(1) The number of events N must be large. [Some people might consider $N > 10$ to be large enough.]

(2) The probability that any member of the initial sample dies or decays (or whatever) must be small. If, for example, we did an experiment to see how many of 100 mice would die within 100 years, the answer would be 100±0. The probability of death is 100%, surely not small. On the other hand, if we start out with 10^8 radioactive nuclei and they decay at the rate of 10^3 per sec, in a 10 sec "experiment" the number of nuclei that decay would be N = 10000 ± 100. The square root rule should work very well because N >>1 and the probability of a given nucleus decaying is 10^{-4} during the experiment.

Estimating Uncertainties

In most of the experiments you will be asked to estimate the error or uncertainty in quantities you measure, for example, a distance or time interval. This requires some common sense, and we might well give some examples. Often the smallest division of a scale or meter will give you a good idea of the uncertainty to apply to a measurement. A reasonable minimum error might then be one-third of the spacing between scale divisions. If the same measurement can be repeated a number of times, you can estimate the error from its reproducibility or by calculating the standard deviation from Eq. 2. If a quantity is determined from the slope of a graph, make reasonable variations of the fitted line.

To estimate the uncertainty in a calculated quantity (one not measured directly), you should use error propagation as described below.

Error Propagation

Usually we cannot make a direct measurement of the quantity we are interested in. We must measure another quantity or quantities and calculate the desired one from them. In our example of the falling ball on the previous page, we might want the acceleration of gravity g. It can be calculated from the fall time t and distance of fall y from $g = 2y/t^2$. We now ask how the standard deviation in g can be obtained if we know the standard deviations of t and y.

The rules for error propagation can be readily derived using calculus. We merely state the results. In this section, we shall call the quantity we want to measure Q and use ΔQ for its uncertainty (or standard deviation) . We use A, B, and C to represent the quantities which are measured directly, such as y and t in our example above, and ΔA, ΔB, . . . are their uncertainties. These quantities can either be means or individual measurements. Note that it is often convenient to work with the fractional error $\Delta Q/Q$.

(1) If Q = cA where c is a constant (with negligible fractional error), then

$$\frac{\Delta Q}{Q} = \frac{\Delta A}{A} \quad \text{or} \quad \Delta Q = c\,\Delta A \tag{5}$$

(2) If $Q = cA^m$ where m is some power, positive or negative (integer or otherwise), then

$$\frac{\Delta Q}{Q} = m\frac{\Delta A}{A} \qquad \text{or} \qquad \Delta Q = cmA^{m-1}\,\Delta A \tag{6}$$

If Q depends on two quantities A and B, the following rules are useful:

(1) If $Q = A + B$ or $Q = A - B$

$$\Delta Q = \sqrt{(\Delta A)^2 + (\Delta B)^2} \tag{7}$$

(2) If $Q = cA^m B^n$ where c is a constant,

$$\frac{\Delta Q}{Q} = \sqrt{\left(\frac{m\Delta A}{A}\right)^2 + \left(\frac{n\Delta B}{B}\right)^2} \tag{8}$$

The general rule for combining terms is given in Eq. 9 below. More often than not, one of the terms will dominate; the others give a small contribution to ΔQ.

With calculators readily available, it becomes practical to calculate the error in Q by "brute force." [This also saves having to remember the above formulas!] Merely calculate Q for the mean value of A, then calculate it again for $A+\Delta A$ (and $A-\Delta A$ if it seems appropriate). The difference will give ΔQ directly. If Q is a function of more than one variable, vary each separately; then combine the separate terms in quadrature, as in Eq. 9 below. Mathematically, if Q is a function of A, B, and C, or $Q = Q(A,B,C)$, the standard deviation in Q is

$$\Delta Q = \sqrt{(\Delta Q)_A^2 + (\Delta Q)_B^2 + (\Delta Q)_C^2} \tag{9}$$

where ΔQ_A is the change in Q when A is varied by one standard deviation, ΔQ_B is the change in Q when B is varied by one standard deviation, *etc.*

As a numerical example, suppose $g = 2y/t^2$ and we know from a series of measurements that $y = (1.010 \pm 0.014)$ meters and $t = 0.454 \pm 0.008$ s. Clearly, the best estimate for g is $g = 2 \text{ x } 1.010/(0.454)^2 = 9.80$ m/s^2 , but what is the uncertainty in g? Using the brute force technique, we find that g changes by approximately 0.36 m/s^2 if we change t by .008 s, and by 0.14 m/s^2 if we change y by .014 m. Then from Eq. 9, the overall uncertainty in g is the square root of the sum of squares of the contributions due to changing t and y,

$$\Delta g = \sqrt{(0.14)^2 + (0.36)^2} = 0.38 \text{ m/s}^2$$

If instead we calculate Δg from Eq. 8 with $g = 2y/t^2$, we have $c = 2$, $A = y = 1.010$, $\Delta A = .014$, $m = 1$, $B = t = 0.454$, $\Delta B = .008$, $n = -2$ and

$$\frac{\Delta g}{g} = \sqrt{\left(\frac{\Delta y}{y}\right)^2 + \left(\frac{-2\Delta t}{t}\right)^2} = \sqrt{\left(\frac{.014}{1.01}\right)^2 + \left(\frac{-2 \text{ x } 0.008}{0.454}\right)^2} = .038$$

Therefore, $\Delta g = .038\, g = .038 \text{ x } 9.80 = 0.37$ m/s^2. Thus either method gives an overall uncertainty in the measurement of g of about 0.38 m/s^2, and we would give our result for g as $g = 9.80 \pm 0.38$ m/s^2. The fractional error in g is $\Delta g/g = 0.038$ or approx. 4%.

Rounding Off and Scientific Notation

The number of significant figures you give in your answer should be consistent with the accuracy of your answer. It doesn't make sense to give too many (*e.g.,* 9.96572 ± .6226) or too few (*e.g.*, 10 ± .0014). Write instead 10.0 ± 0.6 or 9.9563 ±0.0014. Also always use scientific notation for numbers. Don't write A=9813122.0 ± 214667.0; instead use $(9.8 \pm 0.2) \times 10^6$.

Fitting data to a Hypothesis —— Method of Least Squares

Often you will want to know if your data are consistent with a theory and perhaps determine the value of some constant in the theoretical expression which best fits your data. For example, an object dropped from rest is expected to have a velocity at time t

$$v = -g\,t \tag{10}$$

where g is the acceleration of gravity. Eq. 10 is the equation of a straight line, and g could be determined by graphing your data and estimating the slope. Suppose, however, your data were for the position y *vs* time. Then we expect

$$y = y_0 - 1/2\ g\,t^2 \tag{11}$$

This is the equation of a parabola that would be difficult to fit graphically.

There are mathematical techniques for fitting data to polynomials such as Eq. 11 with the general form

$$y = c_o + c_1 t + c_2 t^2 + c_3 t^3 \tag{12}$$

where the c's are constants. These techniques are described in the references given below. It would take us too far afield to go into detail here, but it is worthwhile to explain the general idea and some of the jargon. The usual technique is called the method of least squares. Let y_i be the measured value of y at time t_i and let σ be the standard deviation in y_i. A measure of the goodness of fit of the fitting function (e.g., Eq. 11) to the data is the quantity χ^2 (referred to as "ki squared"),

$$\chi^2 = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} [\, y_i - y(t_i)\,]^2 \tag{13}$$

where N is the number of data points. χ^2 is a measure of the discrepancy between the data points y_i and the fitting function $y(t_i)$, since $[\,(y_i - y(t_i))/\sigma_i\,]$ is just the difference between the measured point y_i and the value from the fitted curve $y(t_i)$ measured in standard deviations. The trick is to find the coefficients c_i in the fitting function, Eq. 12. This is done by choosing them so that χ^2 is minimized. Since χ^2 is a measure of the squares of differences between the data and the curve, the procedure is called the method of least squares.

References on Errors and Treatment of Data

There are two handy paperbacks: Theory of Errors by Yardley Beers (Addison-Wesley Pub. Co.); and Statistical Treatment of Data by Hugh D. Young,(Mc Graw-Hill Pub. Co.). On the lighter side, there is How to Lie with Statistics by Darrel Huff (Norton Pub. Co.). A readable but more advanced book is P.R. Bevington, Data Reduction and Error.