Uncertainties I - An Introduction to Error Analysis

Introduction

Any time a measurement is made there is some uncertainty about the result. Error analysis is the study of that uncertainty. Some knowledge about the uncertainty in a measurement is critical for a proper interpretation of the result. For example, suppose a 5-day old infant showing signs of jaundice has a total serum bilirubin content of 23 mg/dl. The suggested treatment at this level is phototherapy, but if the level is 25 mg/dl or greater, the proper treatment is an exchange blood transfusion. If the doctor knows that the uncertainty in the lab test is less than plus or minus 1 mg/dl, then he or she may feel confident that phototherapy is adequate. However, if the uncertainty is plus or minus 5 mg/dl, the doctor must consider the possibility that the actual level may be well above the level at which a transfusion is recommended.

As another example, consider a measurement designed to test Newton's theory of gravity. A straightforward application of the theory, assuming a spherically symmetric earth, predicts that the acceleration of gravity at the surface of the earth is 9.790 m/s^2. Suppose that a measurement of the acceleration gives the result 9.83 m/s^2. If the uncertainty in the result is 0.01 m/s^2, then clearly the theory and the measurement do not agree, and one must try to ascertain why. On the other hand, if the uncertainty is 0.05 m/s^2, then the experimental result is consistent with the theoretical prediction. In that case one can conclude that any effects due to a non-spherical earth are too small to be observed with the apparatus used for the experiment.

Precision and Accuracy, and Random and Systematic Errors

There are various factors that determine the uncertainty in a measurement. One consideration is the ***precision*** of the measuring device. Precision refers to the size of the smallest increment that can be reliably determined. For example, a mercury thermometer marked in ºC can probably be read reliably to the nearest 0.2 ºC. A digital thermometer that reports temperature to the nearest 0.1 ºC has an inherent precision of 0.05 ºC, or half of the smallest increment reported.

Another consideration is the ***accuracy*** of the measuring device. Accuracy refers to how close the value reported is to the actual value. For example, the electronic balances used in this course report mass to the nearest 0.1 g, but the accuracy is only guaranteed to be good to within 0.02%. If the balance is used to measure a mass of 1000g, then the inaccuracy could be as much as 0.2 g.

In many cases, we obtain an estimate of the uncertainty in a measurement, by repeating the experiment a number of times and examining the variation in the data. For example, suppose that the time for a marble to fall a distance of 1.00 m is measured 6 times with a digital stopwatch that reads to the nearest 0.01 s, and the results are 0.44 s, 0.51 s, 0.45 s, 0.52 s, 0.46 s, and 0.49 s. To obtain the best estimate for the time we take the ***average*** or ***mean*** of the six measurements and get 0.48 s. To estimate the uncertainty we note that the actual value is unlikely to be more than the highest value, 0.52 s, or less than the lowest value, 0.44s. Hence we write the result as

$$t = 0.48 \pm 0.04 \text{ s}.$$

This statement is read t = 0.48 plus or minus 0.04 s, and it is just a shorthand way of saying that the best estimate is 0.48 s and we believe that the actual value is quite likely to be between 0.52 and 0.44 s.

Notice that the uncertainty in this example is much greater than 0.005 s, the value based on the precision of the stopwatch. In this case the uncertainty in the experiment is not determined by the precision of the stopwatch, but is instead determined by other factors such as the reproducibility of dropping the ball, or the ability of the experimenter to start and stop the watch at just the right instant. In this course we will often see that our uncertainty is larger than what one would expect based simply on the precision and accuracy of the measuring instrument.

When we repeat an experiment a number of times and examine the variation in the data, we are obtaining information about the ***random*** errors in the experiment. It is possible that there are also significant ***systematic*** errors in the experiment. Systematic errors are errors that affect all measurements in the same way. For example, perhaps the stopwatch in the above example is designed for operation between 50ºF and 80ºF and when the temperature is above 80º F it runs too fast. If the measurements were made when the temperature was 90ºF, then all of the time readings would be too large. We have no way of knowing about this effect from examining the variation in the data.

Systematic errors can be found (and corrected for) by comparing a result to one known to be more reliable, but often that may not be available, or it may not even exist! Detecting systematic errors and estimating the uncertainty in a measurement due to systematic errors is usually very difficult. In this course we will try to be very cognizant of possible systematic errors, but most of our error analysis will pertain to random errors.

From the above discussion, it should be clear that the precision of a measurement is related to how large the random errors are, while the accuracy of the measurement is related to how large the systematic errors are.

Standard Deviation and the Gaussian Distribution

When a measurement is repeated a large number of times, instead of considering the largest and smallest values to quantify the uncertainty, we calculate σ, the ***standard deviation***. The standard deviation is defined by the equation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}{N-1}},$$

where N is the number of times the experiment is repeated and $\bar{x}$ is the average of the measurements.

To find σ we calculate the difference between each measurement and the average of all measurements, square the difference, add up all the squares, divide by the number of measurements (minus 1!), and then take the square root. In other words, σ is approximately equal to the square root of the average squared difference from the mean. (We will not go into the reasons why N-1 appears in the denominator instead of just N). If all the measurements are the same, then σ = 0. The more variation in the data the larger σ will be. For the six timing measurements discussed above, σ = 0.0331s, a somewhat smaller value than the uncertainty obtained from considering the largest and smallest measurements.

When most measurements are repeated a very large number of times we expect that half of the data will be ≥ the average and half will be ≤ the average. We also expect that most of the results will be fairly close to the average and a smaller number of results will be further from the average (see Fig. 1). The particular distribution of data shown in Fig. 1 is called a ***normal*** or ***Gaussian distribution***. If the data is indeed distributed normally (as we shall usually assume), then it can be shown that about 2/3 of the measurements fall within ± σ of the average, about 95% of the data fall within ± 2 σ of the average, and more than 99% of the data fall within ± 3 σ of the average.

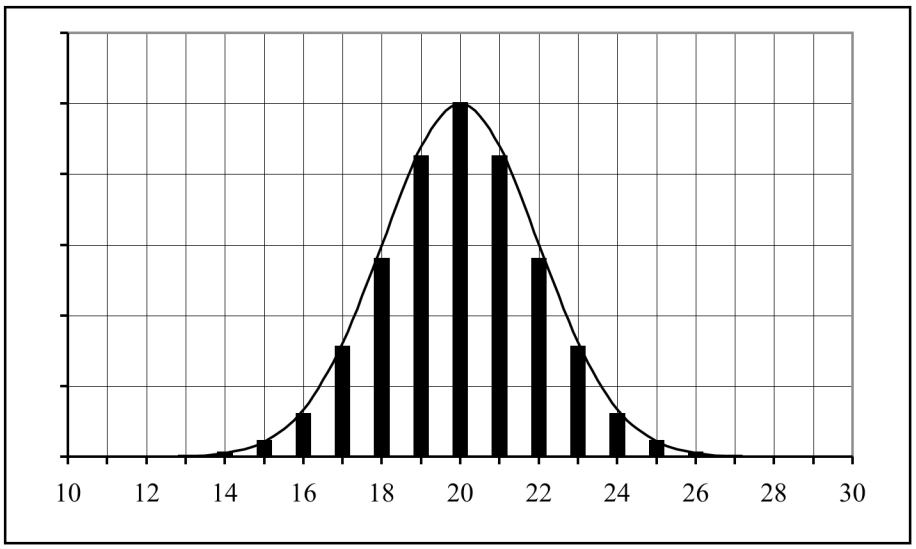

Fig. 1. A histogram of measurements for a *normal* distribution with an average of 20 and a standard deviation of 2. The vertical axis shows the number of occurrences of the value on the horizontal axis.

Reporting Results in Standard Form

In order to save time and effort, we will often make rather crude estimates of the uncertainty in a measurement. To indicate that our uncertainties are only estimates, we normally report the uncertainty with only 1 significant figure. (Even in very careful scientific work, uncertainties are almost never reported with more than 2 significant figures). In the timing example, the standard deviation is actually σ = 0.03311596 s. However, since the measurement was only repeated 6 times the standard deviation calculated is probably not the same as we would get

if the experiment were repeated say, 100 times. As a matter of fact, adding just 1 more measurement of 0.49 s to the data set changes the standard deviation to 0.0305505 s, indicating that only the first significant digit is meaningful. To 1 significant figure the uncertainty is 0.03 s.

When reporting the average of the measurements we should also be aware of how many significant figures we report. In the above example, the actual average of the 6 measurements is 0.47833333 s. However, we report the average to the same precision as the uncertainty. Thus, with an uncertainty of 0.03 s, the result, reported in s***tandard form*** is

$$t = 0.48 \pm 0.03 \text{ s.}$$

Notice that the last digit reported, the 8 in this example, is the digit that is uncertain.

Absolute, Relative and Percent Uncertainty

The uncertainty discussed up to this point is sometimes called the ***absolute uncertainty***, the absolute error, or just the error. When judging the *quality* of a measurement we normally compare the size of the uncertainty to the size of the quantity being measured. For example, measuring the mass of an elephant to the nearest kilogram is quite a good measurement, while measuring the mass of a turkey to the nearest kilogram is not so good. To characterize the quality of a measurement we define the ***relative uncertainty***, as the ratio of the uncertainty to the measurement itself:

$$\text{relative uncertainty} = \frac{\text{absolute uncertainty}}{\text{best estimate}}.$$

Suppose an elephant weighs in at 4681 ± 1 kg. The relative uncertainty is $\frac{1kg}{4681kg} = 0.00021$. If a turkey has a mass of 5 ± 1 kg, the relative uncertainty is $\frac{1kg}{5kg} = 0.2$. The relative uncertainty is just a number because the units cancel.

Usually the uncertainty in a measurement is smaller than the measurement, and it is customary to express the relative uncertainty as a percentage by multiplying by 100. In the case of the elephant, the percentage uncertainty is 0.02% (quite a good measurement) while for the turkey the percentage uncertainty is 20% (not a very good measurement).

Standard Error

Often in this course we will repeat an experiment a number of times and then find the average. Intuitively we know that the more times we repeat the experiment, the more confidence we have in the average value obtained. In other words, the more times we repeat the experiment, the smaller our uncertainty in the average. However, repeating the experiment more times does not have any significant effect on the standard deviation, because the variation of the data is determined by the precision of an individual measurement, and repeating the measurement over and over does not change its precision.

To quantify the decrease in the uncertainty of the average that comes with an increased number of measurements, scientists commonly use the ***standard error***. It is defined by the equation

$$\text{standard error} = \frac{\text{standard deviation}}{\sqrt{N}}$$

where N is the number of times the measurement is repeated.

While the standard deviation indicates the amount of variation of the data about the mean, the standard error expresses how much the mean of N measurements would be expected to vary if the entire N measurements were repeated again.

In the timing example used previously, the standard error is $0.03/\sqrt{6} \cong 0.01\text{s}$. In this case, the standard error indicates that if 6 more time measurements are made, it is quite likely that the average of the 6 new measurements will be within 0.01 s of the previous average of 0.48 s. The standard deviation of the 6 new measurements would still be expected to be about 0.03.

When reporting the result of a measurement with uncertainty, it is always important to specify whether the uncertainty reported is the standard deviation or the standard error.