Chapter 6: THE UNCERTAINTY OF THE MEAN

> " 'Now, for instance, it was reckoned a remarkable thing that at the last party in
>   my rooms, that upon an average we cleared about five pints a head.' "
>   --- *Northanger Abbey*

6.1 USING A SET'S MEAN AS A "BEST GUESS"

Imagine that we are doing an experiment, and we want to make the best possible guess as to the true value of a quantity whose measured values are subject to some kind of random error, either due to human estimation error (as in the case of the stopwatch measurements in the speed of sound lab) or random perturbing physical effects (noise), or both. In order to get a sense of the uncertainty of the measurement, you have already taken N individual measurements of the quantity.

If the effects that perturb the measured value away from its "true value" are truly random, then they are equally likely to cause any given measured value to be *higher* than the true value as they are to cause it to be *lower* than the true value. This means that if you take a set of N measurements of the same quantity and those measurements are subject to random effects, it is likely that the set will contain some values that are too high and some that are too low compared to the true value. If you think of each measurement value in the set as being the sum of the measurement's true value and some random error, what we are saying is that the error of a given measurement is as likely to be positive as negative if it is truly random.

One way to get a good estimate of the true value of a quantity is therefore to calculate the *mean* (average) of the measurement values in the set. When you calculate the mean of a set of measurements, you sum the measurements and then divide by N, the number of measurements in the set. If we imagine each measurement to be equal to the true value plus some random error, then the sum of N measurements will be equal to N times the true value plus the sum of all the random errors. Since the random errors are as likely to be negative as positive, they will tend to cancel each other out when summed, meaning that the sum of N measurements is likely to be very close to N times the true value. Dividing the result by N to get the mean thus yields a number that is likely to be close to the true value. The mean thus represents a *better guess* as to a quantity's true value than any of the individual values in set of measurements.

The assumption that the errors cancel each other out gets better and better as the number of measurements in the set increases. For example, if there are only three measurements in a set, there is one chance in eight that *all three* measurements are too high, implying that the mean of these measurements will *not* be a good estimate of the quantity's true value. As the number of measurements in a set increases, though, the probability that there is an approximately even mix of measured values in the set that are too high and too low becomes larger and larger.

6.2 THE UNCERTAINTY OF THE MEAN

The mean of any finite set of measurements is not going to be exactly equal to the quantity's true value: the random errors are not likely to perfectly cancel (especially if the number of measurements is relatively small. Saying that the mean is a "better guess" than any single given measurement is essentially saying that the uncertainty range of the mean is smaller than the uncertainty range of any given single measurement. But how much smaller is it? How can we quantify the uncertainty of the mean of the set of measurements?

We define the uncertainty of the mean U_m of a set of N measurements to be that value such that we are 95% confident that $\pm U_m$ encloses the "true value" of the measurement (where $\bar{x}$ is the mean of the set). If were to take enough data so that we have many *sets* of N measurements, (as we did in the *Inertial Mass* lab), then this definition means operationally that if U_m has the correct value, the range $\pm U_m$ should enclose the true value for 95% of the sets' means. According to the mathematicians, we can *estimate* the uncertainty U_m of the mean of a normally distributed set of measurement using just *one* set of N measurements as follows:

$$U_m \approx \frac{ts}{\sqrt{N}} \qquad\qquad (6.1)$$

where s is the standard deviation of the set and t is the Student t-factor. In the *Uncertainty of the Mean* lab, you should have verified that this expression was at least approximately consistent with the definition of uncertainty just described.

Note that the estimate of the uncertainty of the mean given by equation 6.1 has two properties that we know (or at least intuitively expect) to be true. First, it implies that the uncertainty of the mean *is indeed* smaller than the uncertainty of a single measurement (by the factor $1/\sqrt{N}$), as we would expect from the argument given in the previous section. Second, the more measurements we take, the smaller U_m becomes, implying that the mean becomes a better estimate of the measurement's true value as the number of measurements it embraces increases (as we should expect).

6.3 ESTIMATING UNCERTAINTIES FROM THE RANGE

Calculating uncertainties of the mean through the standard deviation can quickly become very tedious, particularly when an experiment involves a number of sets of measurements. It would be nice to have a method that allows us to determine the uncertainty of the mean of a set of measurements at a glance.

Here is such a method. The first step is to compute the *range R* of your set of data, which is defined to be the largest value minus the smallest value in your set. (You can often compute R to one or two significant digits in your head.) Then look in Table 6.1 for the number N of the measurements that you took. An *estimate* of the value of the uncertainty (expressed as a multiple of R) appears in the next column, and again, you can probably do the multiplication (to one or two significant digits) in your head. For example, if you have 10 measurements of a period of

time ranging from 2.05 s to 2.22 s, the range is 0.17 s and the uncertainty of the mean of these measurements is (according to the table)

$$U_m = 0.23R = 0.23(0.17 \text{ s}) = 0.04 \text{ s. Easy!}$$

The downside is that the value one reads from this table is only an *estimate* of the value that we would get by calculating $U_m \approx ts/\sqrt{N}$, and not a particularly great one for small N, because it turns out that when N is small, the range of a set of measurements fluctuates from one set to another even more than the calculated value of s does. Still, if we are after a quick and dirty estimate that is still good to about one significant digit, this method is hard to beat.

TABLE 6.1 ESTIMATING U_m FROM THE RANGE			
N	U_m	N	U_m
2	6.40R	10	0.23R
3	1.30R	15	0.15R
4	0.72R	20	0.12R
5	0.51R	25	0.09R
6	0.40R	30	0.07R
7	0.33R	40	0.05R
8	0.29R	50	0.04R
9	0.26R		

We will call this method the **range method** of determining the uncertainty of the mean and the method of directly computing the uncertainty of the mean using $U_m \approx ts/\sqrt{N}$ the **formal method**. You may freely use the range method for quick estimates and for checking formal calculations, but you should be sure to use the formal method when doing the final calculations that you will report in an analysis summary or full report.

6.4 REPEATABLE AND UNREPEATABLE MEASUREMENTS

The discussion in the previous sections suggests that if you want to determine a measured value as precisely as possible, it is better to take a set of measurements of that quantity and compute the mean than it is to take just one measurement. IMPORTANT: In this lab program, when a measurement is repeatable, you should *always* (automatically!) take a *set* of four to twenty measurements of the quantity in question, compute the mean of the set to get a best guess for the measurement value, and then compute the uncertainty of that mean.

But when is a measurement "repeatable"? There are three different kinds of measurements that are *not* repeatable. The first are **intrinsically unrepeatable measurements** of one-time events. For example, imagine that you are timing the duration of a foot-race with a single stopwatch. When the first runner crosses the finish line, you stop the watch and look at the value registered. There is no way to repeat this measurement of this particular race duration: the value that you see on your stopwatch is the only measurement that you have and could ever have of this quantity.

In the second category are **subjectively influenced measurements** where a subjective human judgment is required to make the measurement and such a judgment can be influenced by knowledge of past results or what the answer "should" be. In such cases, while we *can* technically repeat the measurement process, our knowledge of previous results, or our expectation of a certain outcome influences our judgment to such an extent that we cannot be sure that those measurements will remain randomly distributed. For example, imagine measuring

the size of an object using a ruler. The *first* measurement that you take of this quantity may be subject to random effects (for example, the way that you line up the ruler on the object, the orientation of your eye with respect to the ruler, and so on). The second and subsequent measurements that you take of this quantity will be strongly influenced by your knowledge of the first value that you obtained. It is very difficult to be completely unbiased about such measurements after the first, particularly when it comes to reading linear scales; after reading it once, it is hard to avoid looking at the scale the same way subsequently.

This is not generally a problem with digital readouts (where you do not make a subjective judgment about the measurement) or when the random influences overwhelm the influences of subjectivity. For example, if you are using a gauge to read the pressure of steam in a boiler and clearly the reading is fluctuating from moment to moment by substantially more than one division on the scale, your memory of a previous measurement is not likely to have much influence on subsequent measurements.

The final kind of unrepeatable measurement is an **unvarying measurement** whose value always comes out to be exactly the same if we do try to repeat it. This is a particular problem with digital readouts, and is an indication that any random errors that might influence the measurement are smaller than the precision of the instrument. In such a case, equation 6.1 cannot be used because it would clearly yield a zero uncertainty: instead, we have to estimate the uncertainty from our knowledge about the instrument's precision.

Here is a general rule for distinguishing a repeatable from unrepeatable measurements: A measurement is **repeatable** if (1) you are sure that you are really measuring the *same quantity* each time you repeat the measurement, (2) your memory of the earlier measurement will not affect your later measurement, and (3) your measurement device is sufficiently precise that you will see variations in its readout when you repeat a measurement. In general, if you are making a type of measurement that you have never made before, you should repeat the first measurement at least three times: once to estimate the measured value, a second time to see if the measurement varies, and a third time as a check if the first two came out the same. (If the first two *didn't* come out the same, you still want the third measurement because it will improve your statistics.) As you build up your repertoire of measurements, you will learn what sorts of measurements tend to be repeatable and what how large typical uncertainties are for non-repeatable measurements. For example, you have presumably made enough measurements with millimeter-scale rulers to know that you probably will be biased by your first measurement, but that the uncertainty of such measurements is between 0.1 mm and 0.5 mm, depending on your skill at interpolating between the millimeter marks.

6.5 THE BOTTOM LINE

In this laboratory program, you should assign an uncertainty to *every* measured quantity as a matter of course. If the measurement of that quantity is repeatable, take a *set* of measurements (at least three, usually five) and compute the *mean* and the *uncertainty of the mean*. Use the mean as the measurement value in any subsequent calculations.

If the measurement is *not* repeatable, estimate the uncertainty of your single measurement based either on previous experience, your common sense, or on these two principles: (1) the

uncertainty of a linear scale or dial is greater than ± one tenth of the smallest division and (2) the uncertainty of a digital readout is ± 1 in the final digit. For example, you may remember that the uncertainty of stopwatch measurements in the *Inertial Mass* experiment was a certain value: this would be a good estimate of the uncertainty of an unrepeatable stopwatch measurement.

    Remember that all uncertainty estimates (calculated or otherwise) are *estimates* good only to about one significant figure. You will probably typically report calculated uncertainties to two significant figures, to avoid roundoff error. Also check that the reported precision of a quantity's uncertainty is consistent with the reported precision of the quantity itself.

EXERCISES

Exercise 6.1
Which of the following measurements are repeatable? Explain your response.
  (a) Measuring the length of a sheet of paper with a ruler.

  (b) Measuring the width of a building with a meter stick.

  (c) Measuring the outside temperature with a thermometer.

  (d) Measuring how long it takes a rock to drop 5 ft using a watch with a sweep second hand.

  (e) Measuring the duration of a speech with a single watch.

Exercise 6.2
Consider the set of time measurements given in exercise 5.3 in chapter 5 of this lab reference guide. What is the mean of these measurements, and what is its uncertainty? Calculate the latter using both the formal method and the range method. Are the results about the same?