There are three kinds of lies: lies, damned lies, and statistics.

Benjamin Disraeli
(According to Mark Twain)

APPENDIX C:
STATISTICAL MEASURES OF UNCERTAINTY

UNCERTAINTY IS A FACT OF LIFE
In the physical sciences it is often as important to know how certain a given numerical result is as it is to know the result itself. After all, much of science consists of comparing a model world of mathematical theory to the real world that we discover in the laboratory. To make any progress at all, we must have a good idea of just how closely we should expect the two worlds to compare. As a general rule, almost as much work goes into an honest estimate of the uncertainty of a result as went into deriving the result itself. Ideally every number you write down in your lab book should be evaluated as to how certain it is.

SIGNIFICANT FIGURES
In deciding what numbers to report, the concept of significant figures is important. The significant figures in a number are the digits that are obtained directly from the measuring process; zeroes which are included solely for the purpose of locating the decimal point are not considered significant. Let us illustrate this definition with some examples.

<table>
<thead>
<tr>
<th>Number</th>
<th>Number of Significant Figures</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
<td>Implies ±0.5</td>
</tr>
<tr>
<td>7.0</td>
<td>2</td>
<td>Implies ±0.05</td>
</tr>
<tr>
<td>7.00</td>
<td>3</td>
<td>Implies ±0.005</td>
</tr>
<tr>
<td>0.128</td>
<td>3</td>
<td>Leading zero is unnecessary, but it does make the reader notice the decimal point.</td>
</tr>
<tr>
<td>2.324</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2.324 x 10^3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>230</td>
<td>2 or 3</td>
<td>Ambiguous. The zero may be significant or it may be present only to show the location of the decimal point.</td>
</tr>
<tr>
<td>2.30 x 10^3</td>
<td>3</td>
<td>No ambiguity</td>
</tr>
<tr>
<td>2.3 x 10^2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
A measurement and its experimental error should have their last significant digits in the same location (relative to the decimal point).

Examples: \(54.1 \pm 0.1, \ 121 \pm 4, \ 8.764 \pm 0.002.\)

**Handling of Significant Figures in Calculations**

Properly, the correct number of significant figures to which a result should be quoted is obtained via error analysis. However, error analysis takes time, and frequently in actual laboratory practice it is postponed. In such a case, one should retain enough significant figures that round-off error is no danger, but not so many as to constitute a burden. Here is an example:

\[0.77 \times 1.46 = 1.1\ \ \text{WRONG}\]

In this case the numbers 0.77 and 1.46 are known to be accurate to about 1%, whereas the result, 1.1, is accurate to about 10%. In this extreme case, the accuracy of the result is reduced by almost a factor of ten, due to round-off error. Now, a factor of ten in accuracy is usually precious and expensive, and it must not be thrown away by careless data analysis.

\[0.77 \times 1.46 = 1.1242\ \ \text{WRONG}\]

The extra digits, which are not really significant, are just a burden, and in addition they carry the incorrect implication of a result of absurd accuracy.

\[0.77 \times 1.46 = 1.12\ \ \text{okay}\]

\[0.77 \times 1.46 = 1.124\ \ \text{less good, but acceptable}\]

In multiplication or division it is often acceptable to keep the same number of significant figures in the product or quotient as are in the least precise factor. Examples:

\[3.6 \times 25.7 = 92.52 = 92\]

\[4.6 \div 757 = .006077 = .0061\]

Handling of significant figures in addition and subtraction will illustrated with examples:

| \(37.6\) | \(6953\) | \(19.51\) | \(15.43\) |
| \(-2.45\) | \(-42.7\) | \(+.732\) | \(+1.821\) |
| \(35.15\) | \(6910.3\) | \(42.942\) | \(18.5888\) |
| \(35.2\) | \(6910.\) | \(42.9\) | \(18.59\) |

Reporting Results
We usually report a result in the form: best estimate ± uncertainty. Generally, the uncertainty is given to one significant figure (one non-zero number after the decimal point, say), and the least significant figure reported for the best estimate of the result should be of the same order of magnitude as the error. For example, the result of an experiment to measure the distance between two trees might sensibly be reported as 12.8 ± 0.2 m. The best estimate is given to just one decimal place, as is the uncertainty.

A much more precise experiment might yield the result 12.835 ± 0.004 m. The uncertainty is given to one significant figure, but it is in the third place after the decimal point, and so it is reasonable that the best estimate is carried out to three decimal places. It would, however, make no sense to report a result of 12.8 ± 0.2345678 for the first, low precision experiment, nor would it make sense to state the result as 12.843567239 ± 0.234567823!

Don't be misled by all those digits produced by your calculator or by the computer. If you can estimate a velocity to only ±1 cm/sec, don't quote a result like 8.345 cm/sec, even if this is the result you get from, say, an average of a few values. Remember that if you do not state an uncertainty, people will assume an uncertainty of 1 in the last decimal place you quote for a result! The implicit uncertainty in a reported result of 8.345 cm/sec is 0.001 cm/sec.

ESTIMATING THE BEST VALUE:
Generally, one arrives at a best estimate of the quantity under discussion by making a series of measurements and averaging the results. Statisticians will tell you that, as a rule of thumb, you should make at least eight measurements of a quantity before you can 'do statistics' to properly estimate averages and uncertainties! This is rarely convenient, and often impossible to do in the physics laboratory, but whenever you can you should make at least three measurements of any given quantity.

With a minimum of three measures, you stand some chance of spotting problems or out and out blunders. If one of the measurements is radically different from the other two, you can check your results by taking more data. Do not, however, exclude a measurement from the final average unless you are quite sure that the measurement was in error.

SYSTEMATIC ERRORS VS. STATISTICAL UNCERTAINTIES:
There are two kinds of "uncertainties": (1) Systematic errors which increase or decrease all measurements of a quantity in the same sense (either all measurements will tend to be too large, or tend to be too small), and (2) statistical uncertainties which are completely random.
Systematic Errors

Systematic errors usually arise due to inaccuracies in measuring equipment – a spring scale may always read too large a force. But, they can also arise from human errors such as consistently misreading a scale. Systematic errors can also result from the neglect of special conditions. For example, in an experiment to measure a length with a steel ruler, one should take into account the fact that the ruler expands and contracts very slightly with temperature, and is strictly accurate at only one temperature.

Systematic errors are very hard to deal with, but once determined, such systematic effects can be removed from the reported results. For example, one could in principle take a good quality ruler to the National Bureau of Standards and compare it to a very accurate standard metre stick there, and find exactly how much larger or smaller a length it tends to read at some standard temperature. One could then multiply the previous results by the appropriate factor to correct the results.

Sometimes, if you cannot pinpoint the exact effect of a systematic error in ordinary laboratory practice, you can make a reasonable estimate (read “guess”) of how important your systematic errors might be. Then you can correct your data to take this error into account on an approximate basis.

Statistical Uncertainties

Statistical uncertainties arise from a series of small, unknown, and uncontrollable events. As a result, it is impossible for repeated measurements of the same quantity to yield precisely the same value. (If you read a scale or a meter as carefully as you can and are not biased by earlier readings, you will notice small differences in each of your readings.) Often, the best value of a series of readings is given by the average of the readings. Statistical uncertainties are often easier to report than systematic errors, because there are definite mathematical rules for estimating their size.

ESTIMATING UNCERTAINTIES – OVERALL RANGE:

Intuitively, we would expect the uncertainty in a reading to be about the size of the range in readings from the lowest to the highest. In general, repeated measurements of the same phenomenon will not yield the same value. Thus, we usually attribute these differences to random unknown events which change the exact conditions of the measurement slightly from reading to reading. These random events lead to statistical uncertainties. Let's say, for example, that we have a series of measurements of a length: 12.2, 12.2, 12.3, 12.0, 12.1 cm. The overall range in our data would be 12.0 to 12.3 cm, and we would expect that our measurements to be no more uncertain than 0.3 cm.

Statisticians have developed a theory of uncertainty that can put our intuition on a firm footing. There are two mathematical quantities that
are commonly used to describe statistical uncertainties, the standard deviation (or $\sigma$ or S.D. for short), and the standard deviation of the mean (or S.D.M. for short). The S.D.M. is sometimes called "standard error" — a term which we avoid using as the term error in this sense is a misnomer.

When taking measurements in physics we can often assume that variations in measurements are random, or "normally" distributed — just as likely to be too high as too low. If this is the case then both the standard deviation and the standard deviation of the mean can be derived from a set of repeated measurements. The procedure for deriving these terms will be discussed later in this Appendix.

ERRORS ARE NOT UNCERTAINITIES:
There are, of course, true errors, i.e., mistakes. Among these are mistakes in calculations and errors in the use of equipment. Errors can never be completely avoided. It is your responsibility to check and double check calculations and to be sure that equipment is used properly so that errors are minimized.

THE STANDARD DEVIATION:
The standard deviation answers the question "If, after having made a series of measurement and averaging the results, I now make one more measurement, how close is the new measurement likely to come to the previous average?" Here is how you would calculate it if you had a sample of 12 (or more generally $N$) measurements:

1. Sum all the measurements and divide by 12 to get the average or mean.
2. Now, subtract this average from each of the 12 measurements to obtain 12 "residuals".
3. Square each of these 12 residuals and add them all up.
4. Divide this result by ($N$–1) (in this case 11) and take the square root.

Lo and behold you have the standard deviation! We can write out its formula as follows:

Let each of the $N$ measurements be called $x_i$ (where $i = 1$ to $N$) and let the average of the $N$ values of $x_i$ be $\mu$. Then each residual $r_i = x_i - \mu$. Thus:

$$\mu = \frac{x_1 + x_2 + x_3 + x_4 + \ldots + x_N}{N}$$  \hspace{1cm} [C.1]$$

$$r_i = \mu - x_i$$
The standard deviation represents the uncertainty in any one measurement. If you took 50 or 100 measurements, instead of 12, and plotted a histogram of the number of times a given value of \( x_i \) came up against its size you might get a graph that looked like the one shown below.

![Histogram](image.png)

**Figure C-1:** A histogram representing the variation in a set of measurements. The height of each bar is proportional to the number of measurements in each range of values.

This graph is a *histogram*. It represents the number of times a particular value of a measurement, in this case a length, is made for each 0.25 cm interval of length measurement. For example, the histogram represents the fact that 10 of the 100 measurements yielded lengths between 11.950 cm and 11.975 cm. Clearly, values near the middle of the range come up more frequently than do values at either end of the range. If we use the formulas discussed above on the data plotted here, we find that the average or mean lies close to the middle of the range. From this graph one
could estimate a mean of 12.1 cm or so and a probable range of from 12.0 to 12.3 cm.

If, instead of 100 measurements, you took several thousand, your graph could be drawn as a smooth curve, such as the one shown below. For this "bell shaped" curve (also called a "normal distribution" or a "Gaussian") most points lie somewhere between the average value and one standard deviation away from the average value. Since one standard deviation (one “sigma”, \( \sigma \)) in this graph is about 0.1 cm, most points lie between 12.0 and 12.2 cm. In fact, it can be proved that for completely random uncertainties, about 68% of the points will lie within one standard deviation (one \( \sigma \)) from the mean.

![Figure C-2: The curve representing a Gaussian distribution. This is often called the "bell-shaped" curve.](image)

**STANDARD DEVIATION OF THE MEAN (STANDARD ERROR):**

To get a good estimate of some quantity you need several measurements, and you really want to know how uncertain the average of those several measurements is, since it is the average that you will write down (as a best estimate). This uncertainty in the average is what is known as the **standard deviation of the mean** or the **standard error**. Once the standard deviation has been calculated for a sample containing \( N \) measurements, it is very easy to calculate the standard deviation from the mean using the equation:
It is this quantity that answers the question, "If I repeat the entire series of $N$ measurements and get a second average, how close can I expect this second average to come to the first one?" One need not repeat the entire experiment to get the standard error. In fact, a good estimate of the standard error is just the standard deviation of the sample of $N$ measurements divided by the square root of $N$. It is this standard error that one usually reports as the final uncertainty. The abbreviation which is often used for the Standard Deviation of the Mean is S.D.M. Sometimes the S.D.M. is referred to as standard error. Since the S.D.M. is actually a measure of uncertainty rather than of an error, in the sense of a mistake, we prefer not to use this term.

Sample Calculation of Standard Deviation and Standard Deviation of the Mean
Let us imagine that you made the following series of length measurements with a good centimetre ruler: 12.2, 12.2, 12.3, 12.0, 12.1 cm.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>Residual</th>
<th>$(\text{Residual})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.2</td>
<td>+0.04</td>
<td>0.0016</td>
</tr>
<tr>
<td>2</td>
<td>12.2</td>
<td>+0.04</td>
<td>0.0016</td>
</tr>
<tr>
<td>3</td>
<td>12.3</td>
<td>+0.14</td>
<td>0.0196</td>
</tr>
<tr>
<td>4</td>
<td>12.0</td>
<td>+0.16</td>
<td>0.0256</td>
</tr>
<tr>
<td>5</td>
<td>12.1</td>
<td>–0.06</td>
<td>0.0036</td>
</tr>
</tbody>
</table>

$$
60.8 = \frac{60.8}{5} = 12.16 \text{ cm}
$$

Average = 60.8 ÷ 5 = 12.16 cm

Sum of Residuals Squared = 0.0520

\[
\frac{\sum (\text{Residual})^2}{(N-1)} = \frac{0.0520}{4} = 0.0130
\]

Standard Deviation $\sigma = \sqrt{0.0130} = 0.114 \text{ cm}$

Standard Deviation of the mean S.D.M. = $0.114/\sqrt{5} = 0.051 \text{ cm}$

Here the S.D.M. indicates that you are quoting the standard deviation of the mean as your measure of the uncertainty in your best estimate. This represents a 68% confidence level.
THE 95% CONFIDENCE INTERVAL:

Often when you are asked to report data based on measurements, we would like to have you report the mean or average along with a 95% confidence interval with a "±" (plus or minus) sign in front of it. We will use the notation S(95) for this quantity. S(95) is shown on the idealized histogram below. The area under the dark grey part of the probability curve is 95% of the total area under the curve.

![Figure C-3: A Gaussian distribution curve showing the 95% confidence interval.](image)

If you were to take an infinite number of data points, S(95) would simply be twice the standard deviation of the mean (S.D.M).