Appendix 1: Format of Lab Writeups

Lab reports must be between one and two pages, not counting your raw data. For most labs, I will ask you a few questions that your lab writeup should address. The format of your lab report should be as follows:

**Raw data** — *Keep actual observations separate from what you later did with them.*

These are the results of the measurements you take down during the lab, hence they come first. Write your raw data directly in your lab book; don’t write them on scratch paper and recopy them later. Don’t use pencil. The point is to separate facts from opinions, observations from inferences. **You will turn in your raw data (the white carbon-copy) attached to the back of your lab report.**

**Title** — *At the top of your typed lab report.*

**Procedure** — *Did you have to create your own methods for getting some of the raw data?* Do not copy down the procedure from the manual. In this section, you only need to explain any methods you had to come up with on your own, or cases where the methods suggested in the handout didn’t work and you had to do something different. Do not discuss how you did your calculations here, just how you got your raw data. If you followed the lab manual exactly, there’s no need to write a procedure section.

**Abstract** — *What did you find out? Why is it important?*

The “abstract” of a scientific paper is a short paragraph at the top that summarizes the experiment’s results in a few sentences. Many of our labs are comparisons of theory and experiment. The abstract for such a lab needs to say whether you think the experiment was consistent with theory, or not consistent with theory. If your results deviated from the ideal equations, don’t be afraid to say so. After all, this is real life, and many of the equations we learn are only approximations, or are only valid in certain circumstances. However, (1) if you simply mess up, it is your responsibility to realize it in lab and do it again, right; (2) you will never get exact agreement with theory, because measurements are not perfectly exact — the important issue is whether your results agree with theory to roughly within the error bars.

The abstract is not a statement of what you hoped to find out. It’s a statement of what you did find out. Summarize your result.

If this is a lab that has just one important numerical result (or maybe two or three of them), put them in your abstract, with error bars where appropriate. There should normally be no more than three numbers here. Do not recapitulate your raw data; give the summary of the raw data after it has been analyzed, and provide reasoning for the result.

If you’re presenting a final result with error bars, make sure that the number of significant figures is consistent with your error bars. For example, if you write a result as $323.54 \pm 6 \text{ m/s}$, that’s wrong. Your error bars say that you could be off by 6 in the ones’ place, so the 5 in the tenths’ place and the four in the hundredths’ place are completely meaningless.

If you’re presenting a number in scientific notation, with error bars, don’t do it like this

$$1.234 \times 10^{-89} \text{ m/s} \pm 3 \times 10^{-92} \text{ m/s},$$

do it like this

$$(1.234 \pm 0.003) \times 10^{-89} \text{ m/s},$$

so that we can see easily which digit of the result the error bars apply to.

**Calculations and Reasoning** — *Convince me of what you claimed in your abstract.*

Often this section consists of nothing more than the calculations that you started during lab. If those calculations are clear enough to understand, and there is nothing else of interest to explain, then it is not necessary to write up a separate narrative of your analysis here. If you have a long series of similar calculations, you may just show one as a sample. If your prelab involved deriving equations that you will need, repeat them here without the derivation. In some labs, you will need to go into some detail here by giving logical arguments to convince me that the statements you made in the abstract follow logically from your data. For example, your abstract might claim that the gravitational field strength in Fullerton was measured to be

$$g_{\text{exp}} = (9.73 \pm 0.02) \text{ m/s}^2$$

(somewhat inconsistent with the value given in your book) — this section should give the relevant calculations and error analysis that led to that statement. Sound reasoning will get you a good grade; fudging your numbers to get the expected value $(9.80 \pm 0.01) \text{ m/s}^2$ will get you a bad grade.
Appendix 2: Basic Error Analysis

No measurement is perfectly exact.

One of the most common misconceptions about science is that science is “exact.” It is always a struggle to get beginning science students to believe that no measurement is perfectly correct. They tend to think that if a measurement is a little off from the “true” result, it must be because of a mistake — if a pro had done it, it would have been right on the mark. Not true!

What scientists can do is to estimate just how far off they might be. This type of estimate is called an error bar, and is expressed with the ± symbol, read “plus or minus.” For instance, if I measure my dog’s weight to be 52 ± 2 pounds, I am saying that my best estimate of the weight is 52 pounds, and I think I could be off by roughly 2 pounds either way. The term “error bar” comes from the conventional way of representing this range of uncertainty of a measurement on a graph, but the term is also used when no graph is involved.

Some very good scientific work results in measurements that nevertheless have large error bars. For instance, the best measurement of the age of the universe is now* 15 ± 5 billion years. That may not seem like wonderful precision, but the people who did the measurement knew what they were doing. It’s just that the only available techniques for determining the age of the universe are inherently poor.

*UPDATE: The above value for the age of the universe was the measurement as of the early 2000’s. Now (2015) the best guess for the age of the universe is 13.79 ± 0.06 billion years. Because this interval overlaps with the previous interval, we say that the two measurements are consistent. This also shows that cosmology is still a budding science.

Even when the techniques for measurement are very precise, there are still error bars. For instance, electrons act like little magnets, and the strength of a very weak magnet such as an individual electron is customarily measured in units called Bohr magnetons. Even though the magnetic strength of an electron is one of the most precisely measured quantities ever, the best experimental value still has error bars: 1.00115965218091 ± 0.00000000000026 Bohr magnetons. This precision is comparable to measuring the circumference of the Earth to within just under the width of a human hair.

There are several reasons why it is important in scientific work to come up with a numerical estimate of your error bars. If the point of your experiment is to test whether the result comes out as predicted by a theory, you know there will always be some disagreement, even if the theory is absolutely right. You need to know whether the measurement is reasonably consistent with the theory, or whether the discrepancy is too great to be explained by the limitations of the measuring devices.

Another important reason for stating results with error bars is that other people may use your measurement for purposes you could not have anticipated. If they are to use your result intelligently, they need to have some idea of how accurate it was.

Error bars are not absolute limits.

Error bars are not absolute limits. The true value may lie outside the error bars. If I got a better scale I might find that the dog’s weight is 51.3 ± 0.1 pounds, inside my original error bars, but it’s also possible that the better result would be 48.7 ± 0.1 pounds. Since there’s always some chance of being off by a somewhat more than your error bars, or even a lot more than your error bars, there is no point in being extremely conservative in an effort to make absolutely sure the true value lies within your stated range. When a scientist states a measurement with error bars, she is not saying “If the true value is outside this range, I deserve to be drummed out of the profession.” If that was the case, then every scientist would give ridiculously inflated error bars to avoid having her career ended by one fluke out of hundreds of published results. What scientists are communicating to each other with error bars is a typical amount by which they might be off, not an upper limit.

The important thing is therefore to define error bars in a standard way, so that different people’s statements can be compared on the same footing. By convention, it is usually assumed that people estimate their error bars so that about two times out of three, their range will include the true value (or the results of a later, more accurate measurement with an improved technique).
Random and systematic errors.

Suppose you measure the length of a sofa with a tape measure as well as you can, reading it off to the nearest millimeter. If you repeat the measurement again, you will get a different answer. (This is assuming that you don’t allow yourself to be psychologically biased to repeat your previous answer, and that 1 mm is about the limit of how well you can see.) If you kept on repeating the measurement, you might get a list of values that looked like this:

\[
\begin{array}{cccccc}
203.1 & 203.4 & 202.8 & 203.3 & 203.2 \\
203.4 & 203.1 & 202.9 & 202.9 & 203.1 \\
\end{array}
\]

Variations of this type are called random errors, because the result is different every time you do the measurement.

The effects of random errors can be minimized by averaging together many measurements. Some of the measurements included in the average are too high, and some are too low, so the average tends to be better than any individual measurement. The more measurements you average in, the more precise the average is. The average of the above measurements is 203.1 cm. Averaging together many measurements cannot completely eliminate the random errors, but it can reduce them.

On the other hand, what if the tape measure was a little bit stretched out, so that your measurements always tended to come out too low by 0.3 cm? That would be an example of a systematic error. Since the systematic error is the same every time, averaging didn’t help us to get rid of it. You probably had no easy way of finding out exactly the amount of stretching, so you just had to suspect that there might a systematic error due to stretching of the tape measure.

Some scientific writers make a distinction between the terms “accuracy” and “precision.” A precise measurement is one with small random errors, while an accurate measurement is one that is actually close to the true result, having both small random errors and small systematic errors. Personally, I find the distinction is made more clearly with the more memorable terms “random error” and “systematic error.”

The ± sign used with error bars normally implies that random errors are being referred to, since random errors could be either positive or negative, whereas systematic errors would always be in the same direction.

The goal of error analysis

Very seldom does the final result of an experiment come directly off of a clock, ruler, gauge or meter. It is much more common to have raw data consisting of direct measurements, and then calculations based on the raw data that lead to a final result. As an example, if you want to measure your car’s gas mileage, your raw data would be the number of gallons of gas consumed and the number of miles you went. You would then do a calculation, dividing miles by gallons, to get your final result. When you communicate your result to someone else, they are completely uninterested in how accurately you measured the number of miles and how accurately you measured the gallons. They simply want to know
how accurate your final result was. Was it $22 \pm 2$ mi/gal, or $22.137 \pm 0.002$ mi/gal?

Of course the accuracy of the final result is ultimately based on and limited by the accuracy of your raw data. If you are off by 0.2 gallons in your measurement of the amount of gasoline, then that amount of error will have an effect on your final result. We say that the errors in the raw data “propagate” through the calculations. When you are requested to do “error analysis” in a lab writeup, that means that you are to use the techniques explained below to determine the error bars on your final result. There are two sets of techniques you’ll need to learn:

- techniques for finding the accuracy of your raw data
- techniques for using the error bars on your raw data to infer error bars on your final result

### Estimating random errors in raw data

We now examine three possible techniques for estimating random errors in your original measurements, illustrating them with the measurement of the length of the sofa.

#### Method #1: Guess

If you’re measuring the length of the sofa with a metric tape measure, then you can probably make a reasonable guess as to the precision of your measurements. Since the smallest division on the tape measure is one millimeter, and one millimeter is also near the limit of your ability to see, you know you won’t be doing better than $\pm 1$ mm, or 0.1 cm. Making allowances for errors in getting tape measure straight and so on, we might estimate our random errors to be a couple of millimeters.

Guessing is fine sometimes, but there are at least two ways that it can get you in trouble. One is that students sometimes have too much faith in a measuring device just because it looks fancy. They think that a digital balance must be perfectly accurate, since unlike a low-tech balance with sliding weights on it, it comes up with its result without any involvement by the user. That is incorrect. No measurement is perfectly accurate, and if the digital balance only displays an answer that goes down to tenths of a gram, then there is no way the random errors are any smaller than about a tenth of a gram.

Another way to mess up is to try to guess the error bars on a piece of raw data when you really don’t have enough information to make an intelligent estimate. For instance, if you are measuring the range of a rifle, you might shoot it and measure how far the bullet went to the nearest centimeter, concluding that your random errors were only $\pm 1$ cm. In reality, however, its range might vary randomly by fifty meters, depending on all kinds of random factors you don’t know about. In this type of situation, you’re better off using some other method of estimating your random errors.

#### Method #2: Repeated Measurements and the Two-Thirds Rule

If you take repeated measurements of the same thing, then the amount of variation among the numbers can tell you how big the random errors were. This approach has an advantage over guessing your random errors, since it automatically takes into account all the sources of random error, even ones you didn’t know were present.

Roughly speaking, the measurements of the length of the sofa were mostly within a few mm of the average, so that’s about how big the random errors were. But let’s make sure we are stating our error bars according to the convention that the true result will fall within our range of errors about two times out of three. Of course we don’t know the “true” result, but if we sort out our list of measurements in order, we can get a pretty reasonable estimate of our error bars by taking half the range covered by the middle two thirds of the list. Sorting out our list of ten measurements of the sofa, we have

<table>
<thead>
<tr>
<th>Measure</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>202.8 cm</td>
<td>202.9</td>
</tr>
<tr>
<td>202.9</td>
<td>203.1</td>
</tr>
<tr>
<td>203.1</td>
<td>203.2</td>
</tr>
<tr>
<td>203.3</td>
<td>203.4</td>
</tr>
<tr>
<td>203.4</td>
<td></td>
</tr>
</tbody>
</table>

Two thirds of ten is about 6, and the range covered by the middle six measurements is 203.3 cm - 202.9 cm, or 0.4 cm. Half that is 0.2 cm, so we’d estimate our error bars as $\pm 0.2$ cm. The average of the measurements is 203.1 cm, so your result would be stated as 203.1 $\pm 0.2$ cm.

One common mistake when estimating random errors by repeated measurements is to round off all your measurements so that they all come out the same, and then conclude that the error bars were zero. For instance, if we’d done some overenthusiastic rounding of our measurements on the sofa, rounding them all off to the nearest cm, every single number on the list would have been 203 cm. That wouldn’t mean that our random errors were zero! The same can happen with digital instruments that automatically round off for you. A digital balance
might give results rounded off to the nearest tenth of a gram, and you may find that by putting the same object on the balance again and again, you always get the same answer. That doesn’t mean it’s perfectly precise. Its precision is no better than about \( \pm 0.1 \) g.

**Method #3: Repeated Measurements and the Standard Deviation**

The most widely accepted method for measuring error bars is called the standard deviation. Here’s how the method works, using the sofa example again.

1. Take the average of the measurements.
   
   \[
   \text{average} = 203.1 \text{ cm}
   \]

2. Find the difference, or “deviation,” of each measurement from the average.
   
   \[
   -0.3 \text{ cm} \quad -0.2 \quad -0.2 \quad 0.0 \quad 0.0 \\
   0.0 \quad 0.1 \quad 0.1 \quad 0.3 \quad 0.3
   \]

3. Take the square of each deviation.
   
   \[
   0.09 \text{ cm}^2 \quad 0.04 \quad 0.04 \quad 0.00 \quad 0.00 \\
   0.00 \quad 0.01 \quad 0.01 \quad 0.09 \quad 0.09
   \]

4. Average together all the squared deviations.
   
   \[
   \text{average} = 0.04 \text{ cm}^2
   \]

5. Take the square root. This is the standard deviation.
   
   \[
   \text{standard deviation} = 0.2 \text{ cm}
   \]

If we’re using the symbol \( x \) for the length of the couch, then the result for the length of the couch would be stated as \( x = 203.1 \pm 0.2 \text{ cm} \), or \( x = 203.1 \) cm and \( \sigma_x = 0.2 \) cm. Since the Greek letter sigma (\( \sigma \)) is used as a symbol for the standard deviation, a standard deviation is often referred to as “a sigma.”

Step (3) may seem somewhat mysterious. Why not just skip it? Well, if you just went straight from step (2) to step (4), taking a plain old average of the deviations, you would find that the average is zero! The positive and negative deviations always cancel out exactly. Of course, you could just take absolute values instead of squaring the deviations. The main advantage of doing it the way I’ve outlined above are that it is a standard method, so people will know how you got the answer. (Another advantage is that the standard deviation as I’ve described it has certain nice mathematical properties.)

A common mistake when using the standard deviation technique is to take too few measurements. For instance, someone might take only two measurements of the length of the sofa, and get 203.4 cm and 203.4 cm. They would then infer a standard deviation of zero, which would be unrealistically small because the two measurements happened to come out the same.

In the following material, I’ll use the term “standard deviation” as a synonym for “error bar,” but that does not imply that you must always use the standard deviation method rather than the guessing method or the 2/3 rule.

There is a utility on the class’s web page for calculating standard deviations.

### Probability of deviations

You can see that although 0.2 cm is a good figure for the typical size of the deviations of the measurements of the length of the sofa from the average, some of the deviations are bigger and some are smaller. Experience has shown that the following probability estimates tend to hold true for how frequently deviations of various sizes occur:

- < 1 standard deviation about 2 times out of 3
- 1-2 standard deviations about 1 time out of 4
- 2-3 standard deviations about 1 time out of 20
- 3-4 standard deviations about 1 in 500
- 4-5 standard deviations about 1 in 16,000
- > 5 standard deviations about 1 in 1,700,000

The probability of various sizes of deviations, shown graphically. Areas under the bell curve correspond to probabilities. For example, the probability that the measurement will deviate from the truth by less than one standard deviation (\( \pm 1\sigma \)) is about 34 \( \times 2 = 68\% \), or about 2 out of 3. (J. Kemp, P. Strandmark, Wikipedia.)

**Example: How significant?**

In 1999, astronomers Webb et al. claimed to have found
evidence that the strength of electrical forces in the ancient universe, soon after the big bang, was slightly weaker than it is today. If correct, this would be the first example ever discovered in which the laws of physics changed over time. The difference was very small, \(5.7 \pm 1.0\) parts per million, but still highly statistically significant. Dividing, we get \((5.7 - 0)/1.0 = 5.7\) for the number of standard deviations by which their measurement was different from the expected result of zero. Looking at the table above, we see that if the true value really was zero, the chances of this happening would be less than one in a million. In general, five standard deviations (“five sigma”) is considered the gold standard for statistical significance.

However, there is a twist to this story that shows how statistics always have to be taken with a grain of salt. In 2004, Chand et al. redid the measurement by a more precise technique, and found that the change was \(0.6 \pm 0.6\) parts per million. This is only one standard deviation away from the expected value of 0, which should be interpreted as being statistically consistent with zero. If you measure something, and you think you know what the result is supposed to be theoretically, then one standard deviation is the amount you typically expect to be off by — that’s why it’s called the “standard” deviation. Moreover, the Chand result is wildly statistically inconsistent with the Webb result (see the example on page 8), which means that one experiment or the other is a mistake. Most likely Webb at al. underestimated their random errors, or perhaps there were systematic errors in their experiment that they didn’t realize were there.

**Precision of an average**

We decided that the standard deviation of our measurements of the length of the couch was 0.2 cm, i.e., the precision of each individual measurement was about 0.2 cm. But I told you that the average, 203.1 cm, was more precise than any individual measurement. How precise is the average? The answer is that the standard deviation of the average equals

\[
\text{standard deviation of one measurement} \div \sqrt{\text{number of measurements}}.
\]

(An example on page 7 gives the reasoning that leads to the square root.) That means that you can theoretically measure anything to any desired precision, simply by averaging together enough measurements. In reality, no matter how small you make your random error, you can’t get rid of systematic errors by averaging, so after a while it becomes pointless to take any more measurements.
Appendix 3: Propagation of Errors

Propagation of the error from a single variable

In the previous appendix we looked at techniques for estimating the random errors of raw data, but now we need to know how to evaluate the effects of those random errors on a final result calculated from the raw data. For instance, suppose you are given a cube made of some unknown material, and you are asked to determine its density. Density is defined as \( \rho = m/v \) (\( \rho \) is the Greek letter “rho”), and the volume of a cube with edges of length \( b \) is \( v = b^3 \), so the formula

\[
\rho = \frac{m}{b^3}
\]

will give you the density if you measure the cube’s mass and the length of its sides. Suppose you measure the mass very accurately as \( m = 1.658 \pm 0.003 \) g, but you know \( b = 0.85 \pm 0.06 \) cm with only two digits of precision. Your best value for \( \rho \) (with significant figures) is \( 1.658 \pm 0.003 \) g/(0.85 cm)\(^3 \) = 2.7 g/cm\(^3 \).

How can you figure out how precise this value for \( \rho \) is? We’ve already made sure not to keep more than two significant figures for \( m \), but we don’t yet know how uncertain. A simple method for this type of situation is simply to change the raw data by one sigma, recalculate the result, and see how much of a change occurred. In this example, we can add or subtract 0.06 cm to \( b \) for comparison:

\[
\begin{align*}
&b = 0.85 \text{ cm gave } \rho = 2.7 \text{ g/cm}^3 \\
&b = 0.91 \text{ cm gives } \rho = 2.2 \text{ g/cm}^3 \\
&b = 0.79 \text{ cm gives } \rho = 3.4 \text{ g/cm}^3
\end{align*}
\]

The resulting change in the density has magnitude 0.5 or 0.7 g/cm\(^3 \), so we choose the larger error to be an estimate for how much it could have been off by:

\[
\rho = 2.7 \pm 0.7 \text{ g/cm}^3
\]

Propagation of the error from several variables

Note: If you’re in physics 205, 206, 210, or 211, you can skip this and use the above method for estimating error (find the measurement that introduces the biggest error in the calculated value and use that as an estimate for your uncertainty). If you’re in 221-223, you should read this section on propagation of errors.

What about the more general case in which no one piece of raw data is clearly the main source of error? For instance, suppose we get a more accurate measurement of the edge of the cube, \( b = 0.851 \pm 0.001 \) cm. In percentage terms, the accuracies of \( m \) and \( b \) are roughly comparable, so both can cause significant errors in the density. The following more general method can be applied in such cases:

1. Change one of the raw measurements, say \( m \), by one standard deviation, and see by how much the final result, \( \rho \), changes. Use the symbol \( Q_m \) for the absolute value of that change.

\[
m = 1.658 \text{ g gave } \rho = 2.690 \text{ g/cm}^3 \\
m = 1.661 \text{ g gives } \rho = 2.695 \text{ g/cm}^3
\]

\( Q_m = \text{change in } \rho = 0.005 \text{ g/cm}^3 \)

2. Repeat step (1) for the other raw measurements.

\[
b = 0.851 \text{ cm gave } \rho = 2.690 \text{ g/cm}^3 \\
b = 0.852 \text{ cm gives } \rho = 2.681 \text{ g/cm}^3
\]

\( Q_b = \text{change in } \rho = 0.009 \text{ g/cm}^3 \)

3. The error bars on \( \rho \) are given by the formula

\[
\sigma_\rho = \sqrt{Q_m^2 + Q_b^2}
\]

yielding \( \sigma_\rho = 0.01 \text{ g/cm}^3 \). Intuitively, the idea here is that if our result could be off by an amount \( Q_m \) because of an error in \( m \), and by \( Q_b \) because of \( b \), then if the two errors were in the same direction, we might by off by roughly \( |Q_m| + |Q_b| \). However, it’s equally likely that the two errors would be in opposite directions, and at least partially cancel. The expression \( \sqrt{Q_m^2 + Q_b^2} \) gives an answer that’s smaller than \( Q_m + Q_b \), representing the fact that the cancellation might happen.

The final result is \( \rho = 2.69 \pm 0.01 \text{ g/cm}^3 \).

Example: An average

On page 6 I claimed that averaging a bunch of measurements reduces the error bars by the square root of the number of measurements. We can now see that this is a special case of propagation of errors.

For example, suppose Alice measures the circumference \( c \) of a guinea pig’s waist to be 10 cm, Using the
guess method, she estimates that her error bars are about $\pm 1$ cm (worse than the normal normal $\sim 1$ mm error bars for a tape measure, because the guinea pig was squirming). Bob then measures the same thing, and gets 12 cm. The average is computed as

$$c = \frac{A + B}{2},$$

where $A$ is Alice’s measurement, and $B$ is Bob’s, giving 11 cm. If Alice had been off by one standard deviation (1 cm), it would have changed the average by 0.5 cm, so we have $Q_A = 0.5$ cm, and likewise $Q_B = 0.5$ cm. Combining these, we find $\sigma_c = \sqrt{Q_A^2 + Q_B^2} = 0.7$ cm, which is simply $(1.0 \text{ cm})/\sqrt{2}$. The final result is $c = (11.0 \pm 0.7)$ cm. (This violates the usual rule for significant figures, which is that the final result should have no more sig figs than the least precise piece of data that went into the calculation. That’s okay, because the sig fig rules are just a quick and dirty way of doing propagation of errors. We’ve done real propagation of errors in this example, and it turns out that the error is in the first decimal place, so the 0 in that place is entitled to hold its head high as a real sig fig, albeit a relatively imprecise one with an uncertainty of $\pm 0.7$.)

**Example: The difference between two measurements**

In the example on page 5, we saw that two groups of scientists measured the same thing, and the results were $W = 5.7 \pm 1.0$ for Webb et al. and $C = 0.6 \pm 0.6$ for Chand et al. It’s of interest to know whether the difference between their two results is small enough to be explained by random errors, or so big that it couldn’t possibly have happened by chance, indicating that someone messed up. The figure shows each group’s results, with error bars, on the number line. We see that the two sets of error bars don’t overlap with one another, but error bars are not absolute limits, so it’s perfectly possible to have non-overlapping error bars by chance, but the gap between the error bars is very large compared to the error bars themselves, so it looks implausible that the results could be statistically consistent with one another. I’ve tried to suggest this visually with the shading underneath the data-points.

![Error bars and data points](image)

To get a sharper statistical test, we can calculate the difference $d$ between the two results,

$$d = W - C,$$

which is 5.1. Since the operation is simply the subtraction of the two numbers, an error in either input just causes an error in the output that is of the same size. Therefore we have $Q_W = 1.0$ and $Q_C = 0.6$, resulting in $\sigma_d = \sqrt{Q_W^2 + Q_C^2} = 1.2$. We find that the difference between the two results is $d = 5.1 \pm 1.2$, which differs from zero by $5.1/1.2 \approx 4$ standard deviations. Looking at the table on page 5, we see that the chances that $d$ would be this big by chance are extremely small, less than about one in ten thousand. We can conclude to a high level of statistical confidence that the two groups’ measurements are inconsistent with one another, and that one group is simply wrong.
Appendix 4: Graphing

**Review of Graphing**

Many of your analyses will involve making graphs. A graph can be an efficient way of presenting data visually, assuming you include all the information needed by the reader to interpret it. That means labeling the axes and indicating the units in parentheses, as in the example. A title is also helpful. Make sure that distances along the axes correctly represent the differences in the quantity being plotted. In the example, it would not have been correct to space the points evenly in the horizontal direction, because they were not actually measured at equally spaced points in time.

![Graph of motion of a falling object](image)

**Graphing on a Computer**

Making graphs by hand in your lab notebook is fine, but in some cases you may find it saves you time to do graphs on a computer. For computer graphing, I recommend LibreOffice, which is free, open-source software. It’s installed on the computers in rooms 416 and 418. Because LibreOffice is free, you can download it and put it on your own computer at home without paying money. If you already know Excel, it’s very similar — you almost can’t tell it’s a different program.

Here’s a brief rundown on using LibreOffice:

On Windows, go to the Start menu and choose All Programs, LibreOffice, and LibreOffice Calc. On Linux, do Applications, Office, OpenOffice, Spreadsheet.

Type in your x values in the first column, and your y values in the second column. For scientific notation, do, e.g., 5.2e-7 to represent $5.2 \times 10^{-7}$.

Select those two columns using the mouse.

From the Insert menu, do Object:Chart.

When it offers you various styles of graphs to choose from, choose the icon that shows a scatter plot, with dots on it (XY Chart).

Adjust the scales so the actual data on the plot is as big as possible, eliminating wasted space. To do this, double-click on the graph so that it’s surrounded by a gray border. Then do Format, Axis, X Axis or Y Axis, Scale.

If you want error bars on your graph you can either draw them in by hand or put them in a separate column of your spreadsheet and doing Insert, Y Error Bars, Cell Range. Under Parameters, check “Same value for both.” Click on the icon, and then use the mouse in the spreadsheet to select the cells containing the error bars.

**Fitting a Straight Line to a Graph by Hand**

Often in this course you will end up graphing some data points, fitting a straight line through them with a ruler, and extracting the slope.

In this example, panel (a) shows the data, with error bars on each data point. Panel (b) shows a best fit, drawn by eye with a ruler. The slope of this best fit line is 100 cm/s. Note that the slope should be extracted from the line itself, not from two data points. The line is more reliable than any pair of individual data points.

In panel (c), a “worst believable fit” line has been drawn, which is as different in slope as possible from the best fit, while still pretty much staying consistent the data (going through or close to most of the error bars). Its slope is 60 cm/s. We can therefore estimate that the precision of our slope is +40 cm/s.

There is a tendency when drawing a “worst believable fit” line to draw instead an “unbelievably crazy fit” line, as in panel (d). The line in panel (d), with
a very small slope, is just not believable compared to the data — it is several standard deviations away from most of the data points.

### Fitting a Straight Line to a Graph on a Computer

It’s also possible to fit a straight line to a graph using computer software such as LibreOffice.

To do this, right-click on a data-point, and a menu pops up. Choose Insert Trend Line.\(^1\) Choose Linear, and check the box for Show equation.

How accurate is your slope? A method for getting error bars on the slope is to artificially change one of your data points to reflect your estimate of how much it could have been off, and then redo the fit and find the new slope. The change in the slope tells you the error in the slope that results from the error in this data-point. You can then repeat this for the other points and proceed as in appendix 3. In some cases, such as the absolute zero lab and the photoelectric effect lab, it’s very hard to tell how accurate your raw data are \textit{a priori}; in these labs, you can use the typical amount of deviation of the points from the line as an estimate of their accuracy.

### Comparing Theory and Experiment

Figures (e) through (h) are examples of how we would compare theory and experiment on a graph. The convention is that theory is a line and experiment is points; this is because the theory is usually a prediction in the form of an equation, which can in principle be evaluated at infinitely many points, filling in all the gaps. One way to accomplish this with computer software is to graph both theory and experiment as points, but then print out the graph and draw a smooth curve through the theoretical points by hand.

The point here is usually to compare theory and experiment, and arrive at a yes/no answer as to whether they agree. In (e), the theoretical curve goes through the error bars on four out of six of the data points. This is about what we expect statistically, since the probability of being within one standard deviation of the truth is about 2/3 for a standard bell curve. Given these data, we would conclude that theory and experiment agreed.

In graph (f), the points are exactly the same as in (e), but the conclusion is the opposite. The error bars are smaller, too small to explain the observed discrepancies between theory and experiment. The theoretical curve only goes through the error bars on two of the six points, and this is quite a bit less than we would expect statistically.

Graph (g) also shows disagreement between theory and experiment, but now we have a clear systematic

\(^1\)“Trend line” is scientifically illiterate terminology that originates from Microsoft Office, which LibreOffice slavishly copies. If you don’t want to come off as an ignoramus, call it a “fit” or “line of best fit.”
error. In (h), the fifth data point looks like a mistake. Ideally you would notice during lab that something had gone wrong, and go back and check whether you could reproduce the result.
Appendix 5: Finding Power Laws from Data

For many people, it is hard to imagine how scientists originally came up with all the equations that can now be found in textbooks. This appendix explains one method for finding equations to describe data from an experiment.

Linear and nonlinear relationships

When two variables $x$ and $y$ are related by an equation of the form

$$y = cx^p,$$

where $c$ is a constant (does not depend on $x$ or $y$), we say that a linear relationship exists between $x$ and $y$. As an example, a harp has many strings of different lengths which are all of the same thickness and made of the same material. If the mass of a string is $m$ and its length is $L$, then the equation

$$m = cL,$$

will hold, where $c$ is the mass per unit length, with units of kg/m. Many quantities in the physical world are instead related in a nonlinear fashion, i.e., the relationship does not fit the above definition of linearity. For instance, the mass of a steel ball bearing is related to its diameter by an equation of the form

$$m = cd^3,$$

where $c$ is the mass per unit volume, or density, of steel. Doubling the diameter does not double the mass, it increases it by a factor of eight.

Power laws

Both examples above are of the general mathematical form

$$y = cx^p,$$

which is known as a power law. In the case of a linear relationship, $p = 1$. Consider the (made-up) experimental data shown in the table.

<table>
<thead>
<tr>
<th>h=height of rodent at the shoulder (cm)</th>
<th>f=food eaten per day (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>shrew 1</td>
<td>3</td>
</tr>
<tr>
<td>rat 10</td>
<td>300</td>
</tr>
<tr>
<td>capybara 100</td>
<td>30,000</td>
</tr>
</tbody>
</table>

It’s fairly easy to figure out what’s going on just by staring at the numbers a little. Every time you increase the height of the animal by a factor of 10, its food consumption goes up by a factor of 100. This implies that $f$ must be proportional to the square of $h$, or, displaying the proportionality constant $k = 3$ explicitly,

$$f = 3h^2.$$

Use of logarithms

Now we have found $c = 3$ and $p = 2$ by inspection, but that would be much more difficult to do if these weren’t all round numbers. A more generally applicable method to use when you suspect a power-law relationship is to take logarithms of both variables. It doesn’t matter at all what base you use, as long as you use the same base for both variables. Since the data above were increasing by powers of 10, we’ll use logarithms to the base 10, but personally I usually just use natural logs for this kind of thing.

$$\log_{10} h \quad \log_{10} f$$

| shrew 0                      | 0.48 |
| rat 1                       | 2.48 |
| capybara 2                  | 4.48 |

This is a big improvement, because differences are so much simpler to work mentally with than ratios. The difference between each successive value of $h$ is 1, while $f$ increases by 2 units each time. The fact that the logs of the $f$'s increase twice as quickly is the same as saying that $f$ is proportional to the square of $h$.

Log-log plots

Even better, the logarithms can be interpreted visually using a graph, as shown on the next page. The slope of this type of log-log graph gives the power $p$. Although it is also possible to extract the proportionality constant, $c$, from such a graph, the proportionality constant is usually much less interesting than $p$. For instance, we would suspect that if $p = 2$ for rodents, then it might also equal 2 for frogs or ants. Also, $p$ would be the same regardless of what units we used to measure the variables. The constant $c$, however, would be different if we used different units, and would also probably be different for other types of animals.