General Data Analysis Manual
Wilfrid Laurier University

©Terry Sturtevant[1] [2]

January 21, 2009

[1]Physics Lab Supervisor

Contents

List of Figures

List of Tables

Chapter 1

Goals for Science Labs

When you take a science course, you are obviously going to learn some science. However, in a science course *with labs*, you are expected to learn *how to think and act like a scientist*, which is a different thing. It is like the difference between being a spectator for a particular sport and actually being a participant, or like the difference between enjoying cello music and playing the cello. Becoming a professional scientist (like becoming a professional athlete or a professional musician) will require *both* learning science *and* how to think and act like a scientist. In the lectures for science courses, you will learn lots of science. In the labs, you should learn how to think and act as a scientist.

Usually, one thinks of the purpose of a lab as being the *demonstration* or *testing* of laws of some sort, but a lab should accomplish more than that. In fact, learning *how* to demonstrate or test theories or laws is more important (in the lab) than learning the laws themselves.

Learning the *theory* behind any particular experiment can be done in a lecture, and so that is not the main focus of the lab. In fact, for some experiments the "theory" behind them may *never* be covered in class. Part of what you should learn through your university career is how to *assimilate knowledge*. In other words, you are now going to begin to be expected to learn things on your own. In the lab, this may amount to reading the lab manual ahead of time to get an idea of what you will be doing in the lab. (Later in the course there may be more preparation required.) If you feel you need more information, you may need to look it up yourself.

As well, what you learn in the lab should be applicable in other areas of study after you are finished the course. For example, you may never again

need to calculate the acceleration due to gravity, however, if you continue in science you will be sure to do more data analysis of some type.

For these reasons, labs are intended to address three major issues:

1. One has to understand what it *means* to test or demonstrate a law before one can actually do so.
2. To actually perform an experiment, a student will often be required to become familiar with
   (a) lab *apparatus* and *equipment*,
   (b) measurement *techniques* and efficient *methods of gathering data*,
   (c) *methods* and *tools* for *data analysis*.
3. The results of an experiment are only useful if they can be *communicated* to others.

For these reasons, the goals of science *labs* will fall into these general areas:

- develop understanding of *what it means to demonstrate or test laws*
- introduce *measurement tools and equipment*
- introduce *measurement techniques*
- introduce *uncertainty (or error) analysis* tools and techniques
- promote organized presentation of results in the format of formal *lab reports*.

High school labs tend to focus mainly on the *second* of these above; in university courses you should touch on all of them. It should become apparent that the second one is fairly experiment–dependent, whereas the others are less so. While the specific theories tested and the format of the report required may be unique to a particular course, it is important that you as a scientist develop the skills which allow you to investigate theories to assess their validity and communicate your results in a clear and concise manner regardless of your field of study.

Chapter 11

Estimation, Bounding, and Order of Magnitude Calculations

11.1 Estimation

Lots of experiments involve quantities which must be estimated. (For instance, before you *measure* anything, it's good to be able to *estimate* the result you expect, so you can determine what sort of instrument or method you'll need to perform the measurement.) Some estimates may be better than others, but what really matters is that you have a fair idea about how far off your estimate *could* be.

11.1.1 Bounding

Bounding a quantity is forming an estimate of how far off it could be; an *upper bound* is a bound above the expected value, and a *lower bound* is a bound below the expected value.

Picking Realistic Bounds

It's often easy to come up with reasonable bounds for a quantity by using similar known quantities which are pretty clearly above or below. For instance, if you are estimating a person's height, then you can compare with

known heights of family members or friends. If you have to estimate the mass of an object, you can compare it to objects with which you are familiar.

Range of Possible Values for a Quantity

The **range** of values for a quantity is the difference between its upper and lower bounds.

11.1.2 Familiar Comparison

If you're trying to estimate something, and it's similar to something you know, then you can probably make a pretty good estimate by comparing. In other words, if you can establish an upper and a lower bound, then you can estimate something in between.

The goal when making estimates is to try and make them "safe" but "useful"; i.e. you are *pretty sure* about lower and upper bounds on your estimate, but the bounds are close enough together to make the estimate usable.

11.1.3 Less Familiar Comparison

Often it's not easy to make a clear comparison with something very similar, and so the bounds and thus the estimate have to be a bit more fuzzy.

11.1.4 Logarithmic scale

If several people make estimates, they will no doubt vary. However, they will probably still be in a common ballpark. This can be more easily observed by plotting the values on a *logarithmic scale*, such as the one in Figure 11.1. On a logarithmic scale, the distance of a number from the left end of the scale is proportional to the logarithm of the number. Figure 11.2 and Figure 11.3 show some other possibilities. (Logarithmic scales are often identified by the number of **cycles** they show.) A *cycle* is the space between two numbers which differ by a factor of ten. So, between 1 and 10 is one cycle, between 2 and 20 is one cycle, between 5 and 50 is one cycle, etc.
Note that there is *no* zero on a logarithmic scale. All numbers are positive.

Where would zero be, if you wanted to show it?

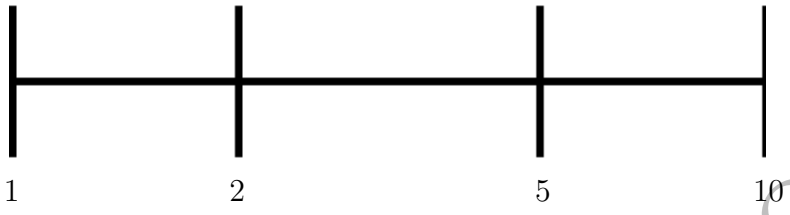


Figure 11.1: Logarithmic scale

There are many things which we *perceive* on a logarithmic scale (such as the volume of music).

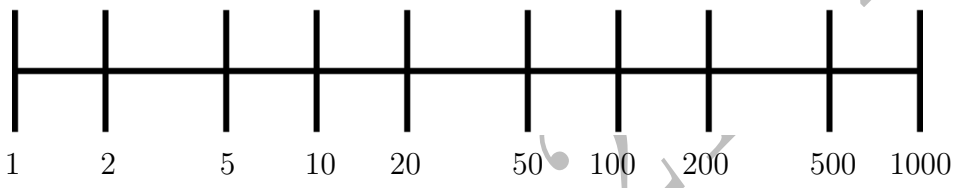


Figure 11.2: Three cycle logarithmic scale

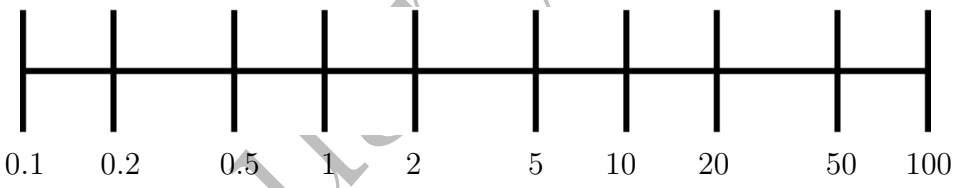


Figure 11.3: Logarithmic scale with numbers less than one

11.2 Order of Magnitude Calculations

More complex quantities can be estimated by performing calculations with estimates. For instance, sometimes certain quantities can be measured but others must be estimated. These calculations are called **order of magnitude calculations**[1], since their purpose is to give a result which is within

[1] or, *"back of the envelope calculations"*

an order of magnitude (i.e. a factor of ten) of the result of the detailed calculation.

How big is an order of magnitude on a logarithmic scale?

If two numbers are within the same cycle of a logarithmic scale, they are within an order of magnitude of each other.

Since an order of magnitude calculation is supposed to be within one order of magnitude, there should be some number, call it K, between 1 and 10 so that $(value \times K)$ is an upper bound and $(value/K)$ is a lower bound. The smaller K is the better. A value of 2 for K means you estimate the correct value to be within a factor of 2 of your calculation; a value of 1.5 for K means you estimate your value to be within 50%, (i.e. a factor of 1.5), of your calculated value, etc.

When is a calculation an Order of Magnitude Calculation?

Any time you have to do a calculation using an estimated quantity, you are performing an **order of magnitude calculation**. The *order of magnitude* of a quantity refers specifically to the power of ten in its measurement. For instance, the height of the building would be in metres, while the length would be in tens of metres. In more general terms, the order of magnitude of a quantity refers to the cycle of a logarithmic scale to which the quantity belongs. Thus we could say that the order of magnitude value for the length of the science building is

- around 100 metres
- between 50 and 200 metres

Both of these are order of magnitude estimates.

When are Order of Magnitude Calculations used?

Order of magnitude calculations are quite commonly done in science *before* an experiment is performed. This is so that the range of expected data can be determined. They are also often done *as* the data are being collected to see if the experimental results appear to be in the correct ballpark.

An order of magnitude calculation is *any* kind of calculation which will produce an answer which should be close to the *"real"* answer. Any calculation involving at least one estimated quantity is an order of magnitude calculation. Generally, the more estimated quantities involved in an order of magnitude calculation, the wider the distance between the upper and lower bounds produced.

11.2.1 Uncertainties

A quantity that is bounded can be expressed as an estimate with an **uncertainty**. (This is a little less cumbersome than giving the estimate, the lower bound, and the upper bound.) Usually it's easiest to express uncertainties in linear (i.e. non-logarithmic) terms, so that an estimate can be given which is *"plus or minus"* some amount. In order to do this, it may require adjusting one of the bounds so that the uncertainty can be the same in both directions. For instance, the length of the building was estimated to be between 50 and 200 metres. If I think it's probably around 100 metres I could modify my estimate of *"between 50 and 200 metres"* to be *"between 50 and 150 metres"* which I could state as "100 ± 50 *metres*".

If you have upper and lower bounds for a quantity, then the uncertainty can be *estimated* as one half of the range; i.e.

$$\textit{uncertainty} \approx 1/2(\textit{upper bound} - \textit{lower bound})$$

(A better determination of the uncertainty will be given in a later exercise.)

Mathematically, the uncertainty in a quantity is usually expressed using the symbol Δ. So in other words, if mass has the symbol m, then the symbol Δm should be interpreted as *"the uncertainty in* m*"*. In that case you would write

$$m \pm \Delta m$$

to mean the mass with its uncertainty. Uncertainty is always given as a positive value, but it can be added or subtracted from the quantity to which it belongs.

11.2.2 Comparing Quantities with Uncertainties

Quantities with uncertainties are said to **agree** if the ranges given by the uncertainties for each overlap. For instance, if I estimated the length of the athletic complex as *"between 60 and 90 metres"* which I could state as *"*75 ± 15 *metres"*, and I estimated the length of the science building as 100 ± 20 metres, then I would say that the lengths of the two building agree since the ranges overlap. In other words, they *may* be the same; without more careful measurement I couldn't say for sure that they are different.

11.3 Recap

By the end of this exercise, you should understand the following terms:

- estimate
- bound
- range of values for a quantity
- logarithmic scale
- order of magnitude calculation
- uncertainty
- whether quantities agree

Chapter 12

Measurement and Uncertainties

> If it's green and wiggles, it's biology;
> If it stinks, it's chemistry;
> If it doesn't work, it's physics.[1]

The quote above is rather cynical, but depending on what is meant by "work", there may be some truth to it. In physics most of the numbers used are not exact but only *approximate.* These approximate numbers arise from two principal sources:

1. *uncertainties* in individual measurements

2. *reproducibility* of successive measurements of the same quantities.

The first of these cases will be discussed in the following section, while the second will be discussed somewhat here, and more later.

12.1 Errors and Uncertainties

When an experiment is performed, every effort is made to ensure that what is being measured is what is *supposed* to be measured. Factors which hinder this are called **experimental errors**, and the existence of these factors results in *uncertainty* in quantities measured.

[1] *The Physics Teacher* 11, 191 (1973)

12.2 Single Measurement Uncertainties

When a number is obtained as a measure of length, area, angle, or other quantity, its reliability depends on the **precision**[2] of the instrument used, the *repeatability* of the measurement, the *care* taken by the experimenter, and on the *subjectivity* of the measurement itself.

12.2.1 Expressing Quantities with Uncertainties

Consider a measured length that is found to be between 14.255 cm and 14.265 cm. A number like this should be recorded as 14.260± 0.005 cm, where the 0.005 cm is the **uncertainty** [3] in the length.

Note: Digits which are not *stated* are definitely uncertain. They are, in fact, unknown, and you can't get any more uncertain than that! For instance, it makes no sense to quote a value of 78.3±0.0003kg. Unless the next three digits after the decimal place are *known* to be zeroes, then the uncertainty due to those unknown digits is much bigger than 0.0003kg. If we actually measured a value of 78.3000kg, then those zeroes should be stated, otherwise our uncertainty is meaningless. (More will be said about significant figures later.)

Remember: The uncertainty in a measurement should always be in the last digit quoted; ie. the least significant digit recorded is the uncertain one.

12.2.2 Random and Systematic Errors

There are two main categories for *errors*, (ie. sources of uncertainty), which can occur: *systematic* and *random*.

- *Systematic* errors are those which, if present, will skew the results in a particular direction, and possibly by a relatively consistent amount. For instance, If we need to calculate the volume of the inside of a tube, and we measure the outer dimensions of the tube, then the volume we calculate will be a little higher than it should be. If we repeated the

[2]This term will be discussed in detail later.

[3]The term *error* is also used for uncertainty, but it suggests the idea of mistakes, and so it will be avoided where possible, except to describe the experimental factors which lead to the uncertainty.

measurement a few times, we might get slightly different results, but they would all be high.

- *Random* errors, on the other hand, cannot be consistently predicted, in direction or size, outside of perhaps broad limits. For example, if you are trying to measure the average diameter of a sample of ball bearings, then if they are randomly chosen there is no reason to assume that the first one measured will be either above average or below.

One of the important differences between random and systematic errors is that systematic errors can be corrected for after the fact, if they can be bounded. (If we measured the thickness of the walls of the tube from the example above, we could use this to correct the volume.) Random errors can only be reduced by repeating the number of measurements. (This will be discussed later.)

It should be noted that a particular measurement may combine both types of errors; if the two above examples are combined, so that one is trying to determine the average inner volume of a bunch of tubes by measuring the outer dimensions, then there would be a systematic error, (due to the difference between inner and outer dimensions), and a random error, (due to the variation between the tubes), which would both affect the results.

12.2.3 Recording Precision with a Measurement

When taking measurements, one can usually estimate a reading to the nearest 1/2 of the smallest division marked on the scale. This quantity is known as the **precision measure** of the instrument. For a *digital* device, you can measure to the least significant digit.

So, for example, if a metre stick has markings every millimetre, then the precision measure is 0.5 mm, and all measurements should be to 10ths of millimetres. On the other hand, if a digital stopwatch measures to 1/100th of a second, the precision measure is 0.01s and measurements should be to hundredths of seconds.

Determine the precision measure of an instrument *before* taking any measurements, not after. Since the number of digits you quote will depend on the precision measure, you cannot make them up after the fact, or assume them to be zero.

12.2.4 Realistic Uncertainties

Sometimes the precision you can actually achieve in a measurement is less than what is theoretically possible. In other words, your uncertainty is not determined by the precision measure of the instrument, but is somewhat *larger* because of other factors.

> The size of the uncertainty you quote should reflect the *real* range of possible values for the quantity measured. You should be prepared to defend any measurement within the uncertainty you give for it, so do not blithely quote the precision measure of the instrument as the uncertainty unless you are convinced that it is appropriate. The precision measure is the *best* that you can do with an instrument; the uncertainty you quote should be what you can *realistically* do. Your goal as you do the experiment is to try and reduce other factors as much as possible so that you can get as close to the precision measure as possible.

There are many possible *sources* for the uncertainty in a single quantity which all contribute to the total uncertainty. The magnitude of each uncertainty contribution can vary, and the uncertainty you quote with the measurement should take all of the sources into account and be realistic. For instance, suppose you measure the length of a table with a metre stick. The uncertainty in the length will come from several sources, including:

- the precision measure of the metre stick
- the unevenness of the ends of the table
- the unevenness of the top of the table (or the side, if you place the metre stick alongside the table to measure)
- the temperature of the room (a metal metre stick will expand or contract)
- the humidity of the room (a wooden table and/or metre stick will swell or shrink)

It is possible to come up with many other sources of uncertainty, but it should be clear that this does not make the uncertainty you use arbitrarily large. In this example, you'd probably ultimately believe your measurement to be within a cm or so, no matter what, and so that is the uncertainty you should

use. (On the other hand, your uncertainty should not be unrealistically small. Even if the metre stick has a precision measure of say, one millimetre, your uncertainty is obviously bigger than that if the ends of the table have variations of 3 or 4 millimetres from one side of the table to the other.)

Usually if the uncertainty in a quantity is bigger than the precision measure, it will be due mainly to a single factor, or perhaps a couple of factors. It is rare that there will be several errors equally contributing to the uncertainty in a single quantity.

Repeatability of Measurements

Whether we repeat a measurement or not, its *realistic* uncertainty should reflect how close we would be able to be if we attempted to repeat the experiment. This reflects many things, including the strictness of our definition of what we are measuring. A later section of the lab manual, Chapter 13, *"Repeated Independent Measurement Uncertainties"*, will discuss how to calculate the uncertainty if we are actually able to repeat a measurement several times.

Here's a guideline for determining the size of the "realistic uncertainty" in a quantity: If someone was to try and repeat your measurement, with only the instructions you have written about how the measurement was made, how big a discrepancy could they reasonably have from the value you got?

Subjectivity

Suppose you are measuring the distance between two dots on a page with a ruler. If the ruler has a precision measure of 0.5mm, but the dots are non-uniform "blobs" which are several millimetres wide, then your *effective* uncertainty is going to be perhaps a few millimeters. The subjectivity in determining the centre of the blobs is responsible for this. When you find yourself in this situation, you should note *why* you must quote a larger uncertainty than might be expected, and *how* you have determined its value.

12.2.5 Zero Error

Some measuring instruments have a certain **zero error** associated with them. This is the *actual* reading of the instrument when the *expected* value would be zero. For instance, if a spring scale reads 5g with no weight hanging on it, then it has a zero error of 5g. Any measurement made will thus be 5g too high, and so 5g must be subtracted from any measurements. (If the zero error was *minus* 5g, then 5g would have to be *added* to every subsequent measurement. Always be sure to check and record the zero error of an instrument with its uncertainty. (Since the zero error is itself a measurement, then it has an uncertainty just like any other measurement.) Subsequent measurements with that instrument should be corrected by adding or subtracting the zero error as appropriate.

With some very sensitive digital instruments, there may be another factor: if the "zero" value of the scale fluctuates over time, then the fluctuation should be taken into account.

12.3 Precision and Accuracy

Two concepts which arise in the discussion of experimental errors are **precision** and **accuracy** which, in general, are not the same thing.

12.3.1 Precision

Precision refers to the number of significant digits and/or decimal places that can be reliably determined with a given instrument or technique. The precision of a quantity is revealed by its uncertainty.
Precision (or uncertainty) can be expressed as either *absolute* or *relative*. In the first case, it will have the same units as the quantity itself; in the latter, it will be given as a proportion or a percentage of the quantity.

Uncertainties may be expressed in the first manner, (i.e. having **units**), are called **absolute uncertainties**. Uncertainties be expressed as a *percentage* of a quantity are then called **percentage uncertainties**.

For example, the measurement of the diameter of two different cylinders with a meter stick may yield the following results:

$$\begin{aligned} d_1 &= 0.10 \pm 0.05\text{cm} \\ d_2 &= 10.00 \pm 0.05\text{cm} \end{aligned}$$

Clearly, both measurements have the same absolute precision of 0.05 cm, i.e., the diameters can be determined reliably to within 0.5 millimeters, but the relative precisions are quite different. For d_1, the relative precision is

$$\frac{0.05}{0.10} \times 100\% = 50.0\%$$

whereas for d_2 it is

$$\frac{0.05}{10.00} \times 100\% = 0.5\%$$

so we could express these two quantities as

$$\begin{aligned} d_1 &= 0.10 \pm 50.0\% \\ d_2 &= 10.00 \pm 0.5\% \end{aligned}$$

An error which amounts to a half a percent in the overall diameter is probably not worth quibbling about, but a fifty percent error is highly significant. Consequently, we would say that the measurement of d_2 is *more precise* than the measurement of d_1. The relative precision tells us immediately that there is something wrong with the first measurement, namely, we are using the wrong instrument. Something more precise is needed, like a micrometer or vernier calipers, where the precision may be more like ± 0.0005 cm.

When comparing quantities, the *more precise* value is the one with the smaller uncertainty.

12.3.2 Accuracy

Accuracy refers to how close the measured value is to the 'true' or correct value. Thus, if a steel bar has been carefully machined so that its length is 10.0000 ± 0.0005 cm, and you determine its length to be 11.00 ± 0.05 cm, your measurement is precise, but inaccurate. On the other hand, if your measurement is 10.0 ± 0.5 cm, it is accurate, but imprecise, and a measurement of the length which yields a value of 10.001 ± 0.005 cm can be considered to be accurate and precise. Errors in precision and errors in accuracy arise from very different causes, as we shall discuss in the next section.

If you do not know what value to "expect" for a quantity, then you cannot determine the accuracy of your result. This will sometimes be the

case. However, even in these cases, you will often still be able to use common sense to determine whether a value is *plausible*. For instance, if you measured the mass of a marble to be 21kg, you should realize that is unreasonable. If you get a value which is unreasonable, you should try and figure out why. (In this case, it may be that the mass should be 21 *grams*; incorrect units are a common source of odd results.)

When comparing quantities, the *more accurate* value is the one which is closer to the "correct" or expected value.

Systematic errors affect the *accuracy* of a measurement or result, while random errors affect the *precision* of a measurement or result.

12.4 Significant Figures

The approximate number 14.26 is said to be *correct to* 4 *figures*, or to have 4 figure accuracy. Those figures that are known with reasonable accuracy are called **significant figures**. *It is permissible to retain only one estimated figure in a result and this figure is also considered significant.*

If three numbers are measured to be 327, 4.02, and 0.00268 respectively, they are *each* said to have three significant digits. Thus, in counting significant figures, the decimal place is disregarded. Zeros at the end of a number are significant *unless they are merely place holders*. If, for example, a mass is found to be 3.20 grams, the zero *is* significant. On the other hand, when the distance to the sun is given to be 150,000,000 km, this is considered to have only 2 significant digits. Note that there is some ambiguity about the significance of the trailing zeros in this case. This can be avoided by the use of **scientific notation**, which for the above measurement would be 1.5×10^8 km. (Note that in this case, zeros are never place holders, and so, if shown, are always significant.) The following rules tell us which digits are significant in an approximate number:

1. all digits other than zero are significant
2. zeros between non–zero digits are significant
3. leading zeros in a number are not significant

4. trailing zeros in a number may or may not be significant. Use the standard form when appropriate to avoid any confusion of this type.

12.4.1 Significant Figures in Numbers with Uncertainties

When quantities have uncertainties, they should be written so that the uncertainty is given to one significant digit, and the the least significant digit of the quantity is the uncertain one. Thus, if a mass is measured to be $152.1g$ with an uncertainty of $3.5g$, then the quantity should be written as

$$152 \pm 4\text{g}$$

(Note the "2" is the uncertain digit.)

> When using scientific notation, you should separate the power of ten from both the quantity and its uncertainty to make it easier to see that this rule has been followed. This is known as the **standard form**. Use the standard form when the quantities you are quoting have placeholder zeroes. When they don't, the standard form is unnecessary and a bit cumbersome. *If you are using the standard form correctly, it should allow you to present results with uncertainties more concisely. Any time that it would be shorter to present a result without using the standard form, it should not be used.*

For instance, if the speed of light was measured to be $2.94 \times 10^8 m/s$ with an uncertainty of $6.3 \times 10^6 m/s$, then it should be written as

$$c = (2.94 \pm 0.06) \times 10^8 m/s$$

Note that this makes the relative uncertainty easier to determine.

12.4.2 Rounding Off Numbers

Often it is necessary to round off numbers. The length 14.26 feet if correct to three figures is 14.3 feet, and if correct to two figures is 14 feet. Following is a list of rules used when rounding off numbers:

1. when the digit immediately to the right of the last digit to be retained is *more than* 5, the last digit retained is increased by one.

2. when the digit immediately to the right of the last digit to be retained is *less than* 5, the last digit retained is unchanged.

3. when the digit immediately to the right of the last digit to be retained *is* 5, the last digit to be retained remains unchanged if even and is increased by one if odd.

This last rule exists so that, for instance, 12.345 rounds to 12.34, but 12.355 rounds to 12.36. If either of these were rounded *again*, they would round correctly. However if the first had been rounded to 12.35, then rounding again would make it 12.4, which is incorrect.

12.5 How to Write Uncertainties

There are different ways of expressing the same uncertainties; which method is used depends on the circumstances. A couple of these will be described below.

12.5.1 Absolute Uncertainty

An **absolute uncertainty** is expressed in the same units as the quantity. Note that uncertainties are always expressed as *positive* quantities. For example, in the quantity

$$123 \pm 4cm$$

"4" is the uncertainty (not "±4"), and both the 123 and the 4 are in cm.

12.5.2 Percentage Uncertainty

An uncertainty can be written as a percentage of a number, so in the above example we could write

$$123 \pm 4 = 123 \pm \left(\frac{4}{123}\right) \times 100\% = 123 \pm 3\%$$

Generally percentage uncertainties are not expressed to more than one or two significant figures.

12.5.3 Relative Uncertainty

Although uncertainties are not actually *presented* this way, they are often *used* this way in calculations, as you will see later. The relative uncertainty is simply the ratio of the uncertainty to the quantity, or the percent uncertainty divided by 100. So again, in the example above, the relative uncertainty is

$$\frac{4}{123}$$

12.6 Bounds on Uncertainty

Occasionally you will have to measure a quantity for which the uncertainty is unknown. In these cases, the uncertainty can be **bounded** by varying the quantity of interest by a small amount and observing the resulting effect. The uncertainty is the amount by which the quantity of interest can be varied with no *measurable* effect. *This is a case where trying to induce more error in an experiment may be desirable!* Suppose that you have a circuit, and you think that the resistance in the wires of the circuit may be affecting your results. You can test this hypothesis by *increasing* the lengths of the wires in your circuit; if your results do not get worse, then there is no evidence that the original length of wires caused a problem.

Chapter 13

Repeated Independent Measurement Uncertainties

If a quantity is measured several times, it is usually desirable to end up with *one* characteristic value for the quantity. (Quoting all the data may be more complete, but will not result in the kind of general conclusion that is desired.) The uncertainty in a quantity measured several times is the *bigger of* the uncertainty in the individual measurements and the uncertainty related to the reproducibility of the measurements. Most of this write-up deals with how to determine the latter quantity.

Sometimes an attempt at a single measurement will give multiple values. For instance, a very sensitive digital balance may produce a reading which fluctuates over time. Recording several values over a period of time will produce a more useful result than simply picking one at random. There are 3 common values which are extracted from data distributions which may be considered "characteristic" in certain circumstances. They are the **mean**, the **median**, and the **mode**. The mean is simply the *average*, with which you are familiar. The median is the "middle" value; *the value which has an equal number of measurements above and below*[1]. (StatsCan often reports the median income, rather than the average. The average wealth of people in Redmond, Washington, where Bill Gates lives, is huge, but it doesn't affect most people.) The mode is the *most commonly occurring value*. (If there is a continuous range of data values, then the data may be grouped into

[1]If there are an even number of points, the median is the average of the two central ones.

smaller "bins" so that a mode of the bins may be defined. Tax brackets for Revenue Canada are an example of these bins.) Depending on the *reason* for the experiment, the choice of a characteristic answer may change.

In a Gaussian, or **normal** distribution, the above decision is simplified by the fact that the mean, median, and mode all have the same value. Thus, *if* the data are expected to fit such a distribution, then an average will probably be a good choice as a quantity characteristic of all of the measurements. The uncertainty in this characteristic number will reflect the distribution of the data. Since the variations in the observations are governed by chance, one may apply the laws of statistics to them and arrive at certain definite conclusions about the magnitude of the uncertainties. No attempt will be made to derive these laws but the ones we need will simply be stated in the following sections.

13.1 Arithmetic Mean (Average)

Note: (In the following sections, each measurement x_i can be assumed to have an uncertainty Δx_i due to measurement uncertainty. How this contributes to the uncertainty in the average, etc. will be explained later.)

The **arithmetic mean** (or average) represents the best value obtainable from a series of observations from "normally" distributed data.

$$\begin{aligned} Arithmetic\ mean &= \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \\ &= \frac{x_1 + x_2 + \cdots + x_n}{n} \end{aligned}$$

13.2 Deviation

The difference between an observation and the average is called the **deviation** and is defined as

$$Deviation = \delta x_i = |x_i - \overline{x}|$$

13.2.1 Average Deviation

The **average deviation**, which is a measure of the uncertainty in the experiment due to reproducibility, is given by

$$Average\ Deviation = \delta = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n}$$

13.2.2 Standard Deviation

The **standard deviation** of a number of measurements is a more common measurement of the uncertainty in an experiment due to reproducibility than the average deviation. The standard deviation is given by

$$\begin{aligned} \textit{Standard Deviation} = \sigma &= \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \\ &= \frac{1}{\sqrt{n-1}}\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} \end{aligned}$$

(One of the main advantages of using the standard deviation instead of the average deviation is that it can be expressed in the second form above which can be simply re–evaluated each time a new observation is made.) With random variations in the measurements, about 2/3 of the measurements should fall within the region given by $\overline{x} \pm \sigma$, and about 95% of the measurements should fall within the region given by $\overline{x} \pm 2\sigma$. (If this is not the case, then either uncertainties were not random or not enough measurements were taken to make this statistically valid.)

This occurs because the value calculated for $\overline{x}$, called the **sample mean**, may not be very close to the "actual" **population mean**, μ, which one would get by taking an infinite number of measurements. (For example, if you take 2 measurements of a quantity and get values of 1 and 2 respectively, should you guess that the "actual" value is 1, 1.5, 2, or something else?) Because of this, there is an uncertainty in the calculated mean due to the random variation in the data values. This uncertainty will be discussed further in the following section.

Rule of thumb: For normally distributed data, an order of magnitude approximation for the standard deviation is 1/4 the range of the data. (In other words, take the difference between the maximum and minimum values and divide by 4 to get an approximate value for the standard deviation.)

13.3 Standard Deviation of the Mean

(In some texts this quantity is called the "standard *error* of the mean".) Once a number of measurements have been taken, and a mean calculated,

one may calculate the uncertainty in the calculated mean *due to the scatter of the data points*, (ie. reproducibility). Or to be more precise, one can calculate an *interval* around the *calculated* mean, $\overline{x}$, in which the *population* mean, μ, can be reasonably assumed to be found. This region is given by the *standard deviation of the mean*,

$$\textit{Standard deviation of the mean} = \alpha = \frac{\sigma}{\sqrt{n}}$$

and one can give the value of the measured quantity as $\overline{x} \pm \alpha$. (In other words, μ should fall within the range of $\overline{x} \pm \alpha$.)

If possible, when doing an experiment, enough measurements of a quantity should be taken so that the uncertainty in the measurement due to instrumental precision *is greater than or equal to α. This is so that the random variations in data values at some point become less significant than the instrument precision. (In practice this may require a number of data values to be taken which is simply not reasonable, but sometimes this condition will not be too difficult to achieve.)*

In any case, the uncertainty used in subsequent calculations should be the *greater* of the uncertainty of the individual measurements and α.

In mathematical terms[2],

$$\Delta\overline{x} = max\,(\alpha, \text{p.m.})$$

since p.m., the precision measure of the instrument, would be the uncertainty in the average due to the measurement uncertainties alone.

(Note that you need not calculate uncertainties when calculating the average deviation, the standard deviation, and the standard deviation of the mean, since these quantities are used to determine *the uncertainty in the data due to random variations.)*

[2] This is only strictly true if the precision measure is the uncertainty in each of the individual measurements. It is possible that there would be different uncertainties in different measurements, in which case the result should be written $\Delta\overline{x} = max\left(\alpha, \overline{(\Delta x_i)}\right)$, where Δx_i is the uncertainty in measurement i.

i	x_i	x_i^2
1	1.1	1.21
2	1.4	1.96
3	1.3	1.69
4	1.2	1.44
n	$\sum x_i$	$\sum x_i^2$
4	5.0	6.3

Table 13.1: Sample Data

13.4 Preferred Number of Repetitions

Since the the uncertainty of the average is the *greater* of the uncertainty of the individual measurements and α, and α decreases with each additional measurement, then there is a point at which *alpha* will equal the precision measure. In this case, the experiment is "optimized" in the sense that in order to improve it (ie. reduce the uncertainty in the result), one would have to get a more precise instrument *and* take more measurements. This situation occurs when

$$\alpha = \text{p.m.}$$

so if we set

$$\text{p.m.} = \frac{\sigma}{\sqrt{n}}$$

and solve for n, then the result will be the optimal number of repetitions. (Keep in mind that σ does not change much after a few measurements, so it can be calculated and used in this equation.)

13.5 Sample Calculations

Following is an example of how the mean, standard deviation, and standard deviation of the mean are calculated. (The x_i values represent a set of data; x_1 is the first value, x_2 is the second, etc.)

Therefore

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5.0}{4} = 1.25$$

and so

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}} \\ &= \frac{1}{\sqrt{n-1}}\sqrt{\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}} \\ &= \frac{1}{\sqrt{4-1}}\sqrt{(6.3)-\frac{(5.0)^2}{4}} \\ &= 0.129099\end{aligned}$$

thus

$$\alpha = \frac{\sigma}{\sqrt{n}} = \frac{0.129099}{\sqrt{4}} = 0.06455$$

The uncertainty which should be quoted with the average above will be the *bigger* of the uncertainties in the individual measurements and the standard deviation of the mean. So, if the above x_i values all had an uncertainty of 0.05, then since 0.05 is less than α, we would write

$$\overline{x} = 1.25 \pm 0.06$$

If, on the other hand, the x_i had an uncertainty of 0.07 units, then we would write

$$\overline{x} = 1.25 \pm 0.07$$

since 0.07 is greater than α.

Note that in both of these cases, the uncertainties have been rounded to one significant digit, and the average is rounded so that its last significant digit is the uncertain one, as required.

13.6 Simple Method; The Method of Quartiles

There is a way to get values very close to those given by calculating the mean and standard deviation of the mean with very little calculation. (This will be true if the data have a Gaussian[3] distribution.) The method involves

[3] or "normal"

dividing the data into *quartiles*. The first quartile is the value which is above 1/4 of the data values; the second quartile is the value which is above 1/2 of the data values[4] and so on. The second quartile gives a good estimate for the average, and the third quartile minus the first quartile gives a good estimate[5] for the standard deviation. Thus,

$$\overline{x} \pm \alpha \approx Q_2 \pm \frac{(Q_3 - Q_1)}{\sqrt{n}}$$

If you use a number of data values which is a perfect square, such as 16, then the only calculation is *one* division!

13.7 Recap

When dealing with a set of numbers, calculate

- the average, $\bar{x}$
- the standard deviation, σ
- the standard deviation *of the mean*, α
- the uncertainty in the average, which is the bigger of the precision measure of the instrument and the the standard deviation of the mean

The range of the set of numbers (biggest minus smallest) is about four times the standard deviation.

[4] which is also the median

[5] Actually the inter-quartile distance or $IQR \approx 1.35\ \sigma$ for normally distributed data.

Chapter 14

Repeated Dependent Measurements: Constant Intervals

14.1 The Method of Differences

Occasionally, one performs an experiment which produces *more* data than is absolutely necessary. (This is especially true when using electronic equipment which automates data collection.) For example, when doing an experiment measuring the position of a cart on a track at fixed time intervals you may produce several data points as shown in Figure 14.1.

Consider the situation in Figure 14.1. The first 3 cases all have 6 dots spaced 1 second apart, for a total time interval of 5 seconds. The fourth case has only two dots, but they are spaced 5 seconds apart, so again the total time interval is 5 seconds. If we were to calculate the velocity using endpoints alone, all four sets of data would give the same answer. The important question for us is *"Which data set do you trust the most?"*, or in more scientific terms, *"Which data set has the smallest uncertainty?"*

1. Consider cases 1 and 2. The centres of the corresponding dots are at exactly the same times, so each sub-interval is identical. However, the dots in case 1 are much larger. Because of the *dot size*, we would prefer the data in case 2, or in other words, case 2 has the smaller uncertainty due to the smaller dots.

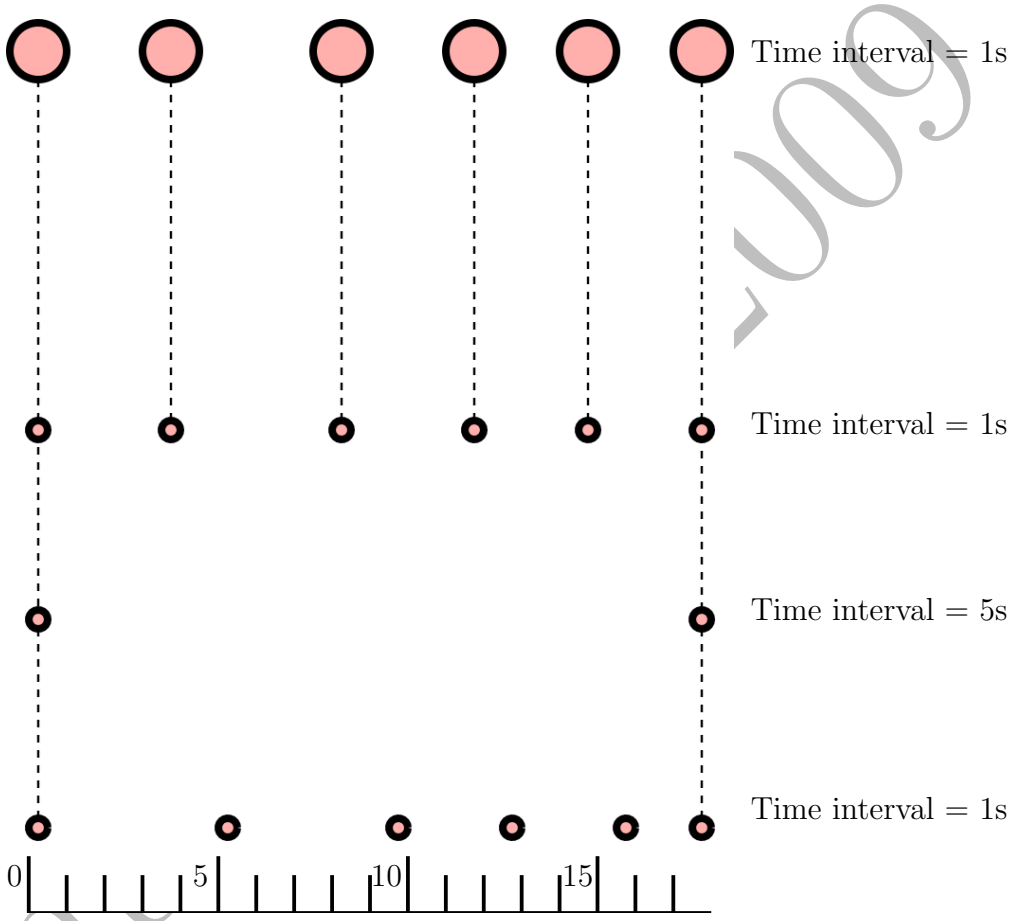


Figure 14.1: Different Possibilities

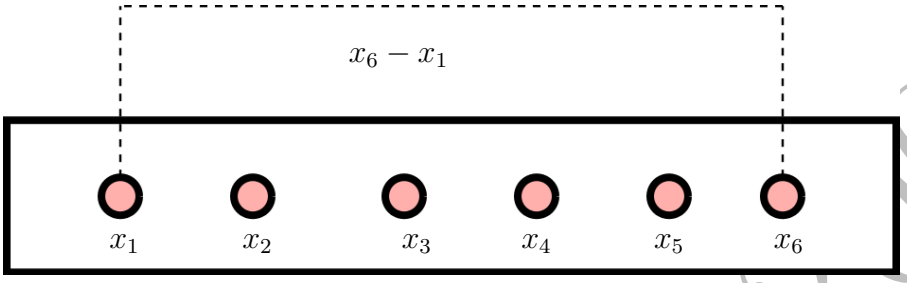


Figure 14.2: Position Measurements for Constant Velocity

2. Consider cases 2 and 3. The dot sizes are the same for both cases. What about variations? Since case 3 has only *1* sub-interval, we have no idea how consistent that measurement would be with a subsequent one. Because of this, case 2 would be preferable because it gives us uncertainties due to both factors (dot size and interval variations), whereas case 3 only addresses one of them.

3. Consider cases 2 and 4. The dots are the same size in both of these cases, however, the widths of the sub-intervals in case 4 vary much more than those of case 2. (In fact, it is clear in case 4 that the object is decelerating.) In this situation, we would say that case 2 has a smaller uncertainty than case 4 due to the *variations in the dot spacings*.

Question: If your choice was only between cases 1 and 4, how would your uncertainties make it possible for you to choose which is better?

If velocity is approximately constant, we could calculate it with just 2 points, even though we have many more. Making use of the "extra" data is what this section is all about. If you need to have at least two data points to perform a calculation, then the data are "dependent".

If the points are approximately *equally spaced*, (i.e. velocity is approximately constant), the velocity can be calculated as follows. (If the velocity is *not* approximately constant, the analysis is more complex and will be discussed later.)

Suppose, for instance, you have 6 marks, which are measured to give positions x_1 to x_6 at times t_1 to t_6, as in Figure 14.2.

14.1.1 Option I: End Points

The simplest way we can calculate $\overline{v}$ is to use just two points. Clearly, the uncertainty in $\overline{v}$ will be minimized if we use points that are as far apart as possible to calculate the interval width; i.e.

$$\overline{v} = \frac{x_6 - x_1}{t_6 - t_1} \tag{14.1}$$

The uncertainty in the result due to the precision measure of the measuring instrument can be calculated as usual. Now, is there any way we can use *all* of the data to produce better results?

14.1.2 Option II (Better):Method of Differences

While the data can be considered as 1 interval of width 5, we can use the points we ignored and see the data as 3 independent intervals of width 3, as in Figure 14.3.

Suppose we calculate $\overline{x}$ as

$$width_1 = x_6 - x_3 \tag{14.2}$$

$$width_2 = x_5 - x_2 \tag{14.3}$$

$$width_3 = x_4 - x_1 \tag{14.4}$$

and then

$$\overline{x} = \frac{width_1 + width_2 + width_3}{3} \tag{14.5}$$

(the average of the 3 widths) and similar for $\overline{t}$. The velocity will be given by

$$\overline{v} = \frac{\overline{x}}{\overline{t}} \tag{14.6}$$

In this way, we turn 1 interval of 5 into *3* independent intervals of 3! Now these 3 independent measurements can be treated as usual, with standard deviations, etc. The value in this is that now all of the data are used, and variations in the interval spacing will affect the uncertainty by increasing the standard deviation of the mean. (Note each measurement is used once and only once, so all have equal weight.)

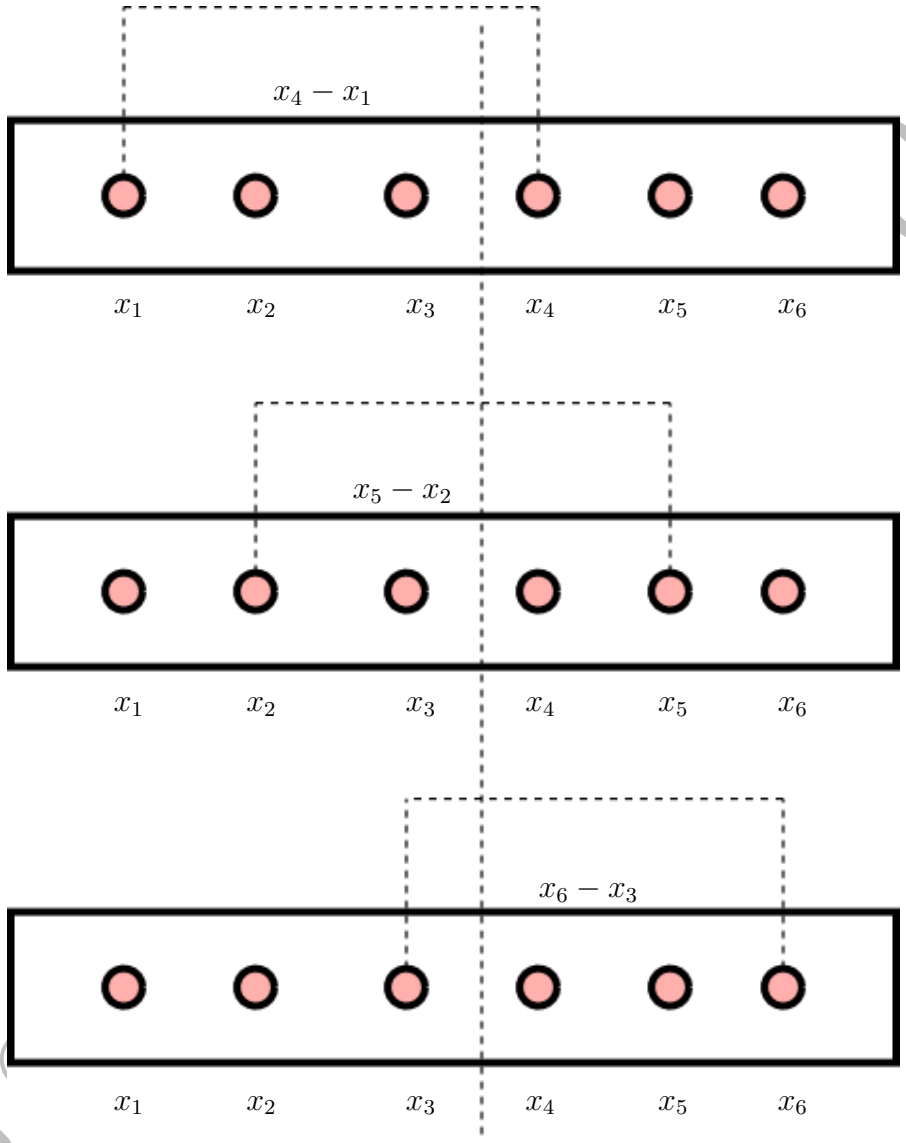


Figure 14.3: Independent Sub-intervals

The uncertainty in your result will be the bigger of the uncertainty from the precision measure and the standard deviation of the mean.(In this example, the uncertainty due to the precision measure would be $2\Delta x$. If this is unclear, ask.)

Whenever you have data like this, you should use this method.

Method of Differences with an Odd Number of Points

To see how this method works with an odd number of points, see Figure 14.4. In this case we can ignore one value and use the remaining (even) number as before. *Note that to keep all of the intervals the same, there are only three choices of which point to drop; the first, the last, or the middle.*

14.1.3 Non-Option III :Average All of the Sub-Intervals

You may be wondering why we don't average *all* of the sub-intervals; in other words, why not determine

$$\overline{v} = \overline{\frac{x_{i+1} - x_i}{t_{i+1} - t_i}} \tag{14.7}$$

for $i = 1, 2, \ldots 6$. If you expand this out, you get

$$\overline{v} = \frac{\frac{x_6 - x_5}{t_6 - t_5} + \frac{x_5 - x_4}{t_5 - t_4} \cdots \frac{x_3 - x_2}{t_3 - t_2} + \frac{x_2 - x_1}{t_2 - t_1}}{5} \tag{14.8}$$

Since all of the time intervals are the same, this becomes

$$\overline{v} = \frac{(x_6 - x_5) + (x_5 - x_4) \cdots (x_3 - x_2) + (x_2 - x_1)}{5\Delta t} \tag{14.9}$$

Removing the brackets gives

$$\overline{v} = \frac{x_6 - \cancel{x_5} + \cancel{x_5} - \cancel{x_4} \cdots \cancel{x_3} - \cancel{x_2} + \cancel{x_2} - x_1}{5\Delta t} \tag{14.10}$$

Notice we have both a '+' and a '-' term for each of x_2 to x_5, so that we are left with

$$\overline{v} = \frac{x_6 - x_1}{5\Delta t} \tag{14.11}$$

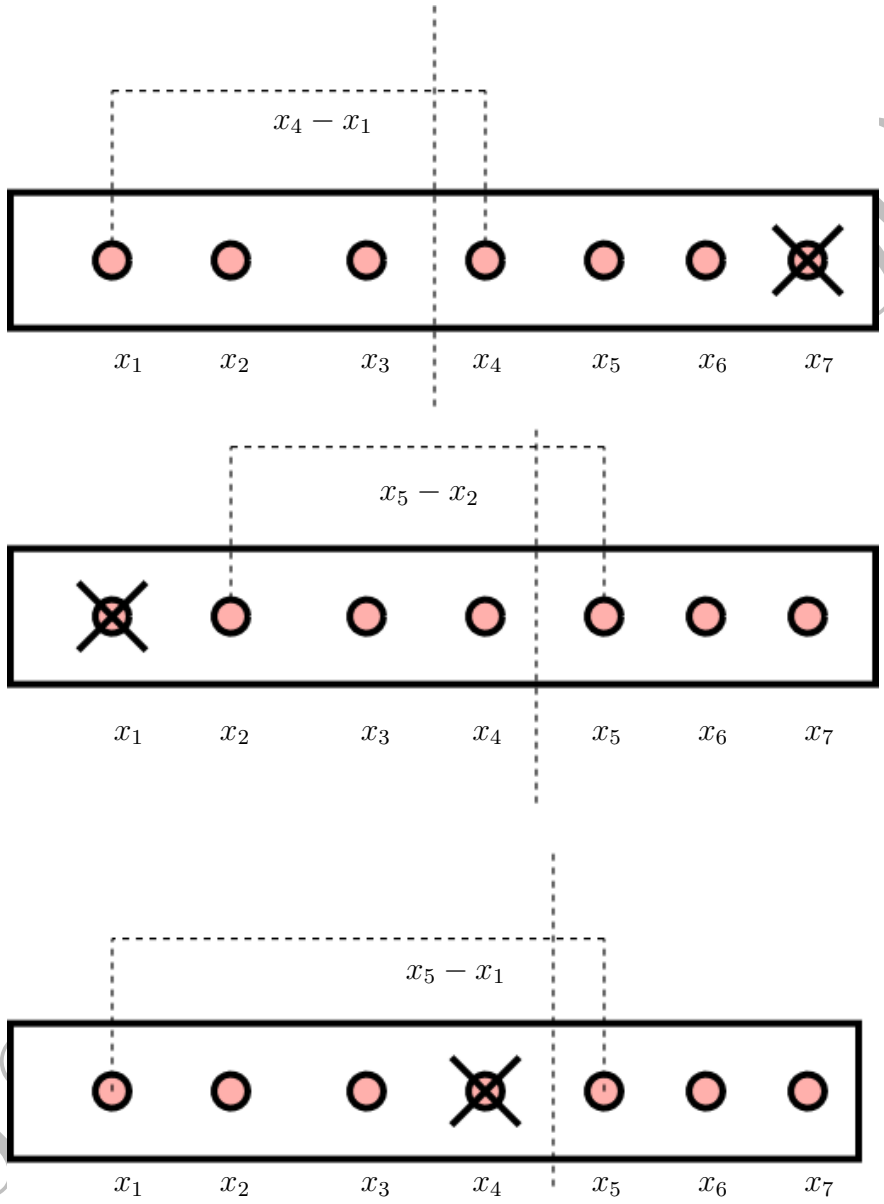


Figure 14.4: Working with an odd number of points

which is exactly what we get by using the endpoints alone! This is why, *with equal time intervals*[1], there is no point to averaging all of the intervals.

14.2 Recap

When dealing with a set of dependent values, do the following:

- Divide the set in half.
- Take differences by subtracting the first value in one half from the first value in the other; the second value in one half from the second in the other, etc.
- Treat the *differences* like independent measurements. (i.e. calculate the average, standard deviation, etc. of the differences.)

(If you have an *odd* number of values, omit either the first, last, or middle value.)

[1]Note that if the time intervals are not the same, then the simplification from Equation 14.8 to Equation 14.9 cannot happen, so the intermediate points do not automatically disappear.

Chapter 15

Repeated Dependent Measurements: Varying Intervals

15.1 Introduction

Previously, it was discussed how to deal with repeated dependent measurements where the intervals varied randomly. The case which will now be addressed is the one in which the intervals are changing *monotonically*, either increasing or decreasing. Without knowing or suspecting some pattern to the variations, one cannot easily determine how to interpret the data. However, if the relationship between the spacings is known, then analysis can be done. The following discussion deals with one special case.

If a cart on a track travels at a speed which is roughly constant, the velocity of the cart can be best calculated using the **method of differences**, as discussed previously. If, however, the cart is accelerating or decelerating, then the procedure for interpreting the results is a little more complex. If the acceleration (or deceleration) is expected to be *non-uniform*, then it is necessary to have some idea of the nature of the relationship expected before proceeding. The simplest case of *uniform* acceleration (or deceleration) can be dealt with as follows.

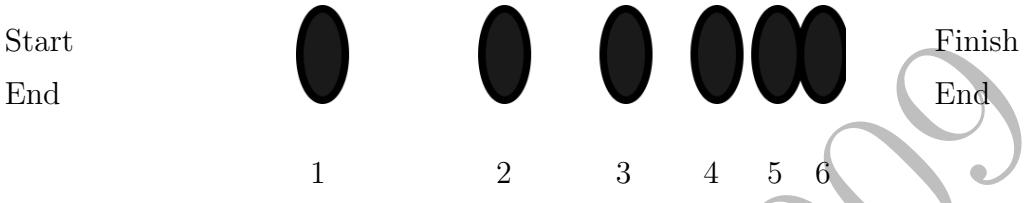


Figure 15.1: Spark Tape for Varying Velocity

15.2 Special Case: Constant Acceleration

(Note that the example speaks of a spark tape with *distance* marks at equal *time* intervals, but clearly the analysis is similar if instead we have fixed *distance* measurements at varying *time* intervals.)

Consider the spark tape shown in Figure 15.1.

A body traveling with constant acceleration obeys the equations of motion which follow:

$$v = v_0 + at \tag{15.1}$$

$$x = x_0 + \overline{v}t \tag{15.2}$$

$$x = x_0 + v_0 t + \frac{1}{2}at^2 \tag{15.3}$$

where $\overline{v}$ is the *average velocity*. If we choose to start our measurements at one of the marked points on the tape, (i.e. $x_0 = 0$), then Equation 15.3 above reduces to

$$x = v_0 t + \frac{1}{2}at^2 \tag{15.4}$$

Dividing both sides of this equation by t gives

$$\frac{x}{t} = v_0 + \frac{1}{2}at \tag{15.5}$$

Equation 15.5 is the equation of a straight line. Thus, if we calculate x/t for each data point and plot x/t vs. t for the data, we should get a straight line with a slope of $\frac{1}{2}a$ and a y intercept of v_0. (Note for the first point, since x and t are both zero, then x/t is undefined so we can't use it.) The velocity at any point is then given by Equation 15.2. (Whether or not the acceleration *actually* was constant will be shown by whether or not the graph was linear.)

One of the nice features of using this approach is that it doesn't matter whether the object was at rest initially. Often, this can't be controlled very well and so it is nice to be able to get equally accurate results either way. As previously stated, the first point (i.e. the starting point) will not give a data point since $x = 0$ and $t = 0$ and thus x/t is undefined, and so can't be used for graphing or for calculations. (The advantage gained by having a linear graph outweighs the loss of one data point.)

15.3 General Case: Non-Uniform Acceleration

As stated above, in the general case, to summarize the results it will be necessary to fit the data to some mathematical relationship, but that is beyond the scope of this course.

Chapter 16

Measuring Instruments and Uncertanties

16.1 Theory

Doing experiments in science involves measuring. In order to take useful measurements, a few concepts need to be understood:

16.1.1 Range

Any instrument has a limit to the values it can measure. For instance, a metre stick can only measure lengths up to a metre. When you choose an instrument to measure something, you are probably making an estimate in your head of how big the things is you're going to measure to be sure that the instrument you use will work.

Examples of range

What is the largest measurement you can make with each of these?

1. Vernier caliper (approximately)
2. micrometer caliper (approximately)
3. spring scale A
4. spring scale B

Why have multiple instruments for one quantity?

Since not all instruments measuring one quantity, (such as length), have the same range, why wouldn't you just use the instrument with the biggest range for all measurements? (e.g. Why wouldn't you just use a meter stick for all length measurements?)

16.1.2 Precision

If you want to compare two objects to see if they are the same in some way, such as whether two marbles have the same mass, you need to measure them with some instrument. The **precision** of an instrument refers to *how close two measurements can be and still be distinguished.* Usually instruments with a large range don't have as much precision, (or, "are not as precise"), as instruments with a small range.

Digital instruments

If you have a digital clock, which shows hours and minutes, how close can two times be and still be different? Obviously, if they are at least 1 minute apart, then they are different. What about a stopwatch that measures to hundredths of seconds? Times that are at least one hundredth of a second apart will be distinct. Since these times are much closer than the times which the digital clock can distinguish, we say the stopwatch is *more precise* than the clock.

We call this *smallest difference between two measurements which can be distinguished* the **precision measure** of an instrument. A smaller precision measure indicates a more precise instrument.

Scales and analog instruments

Many measuring instruments are not digital; they are **analog**. This means that rather than giving an unambiguous disctinct value, they show a continuous range of possible values. Consider the scale reading in Figure 16.1 below:

First of all, the left edge of the object is lined up with the zero of the scale, so we should be able to read the length of the object from the scale at the right edge of the object. It's pretty clear that that the object ends between the '7' and the '8' of the scale. Now take a look at another object

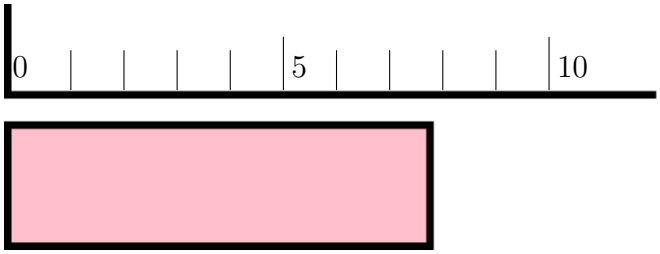


Figure 16.1: An analog scale

measured with the same scale in Figure 16.2. This object also ends between

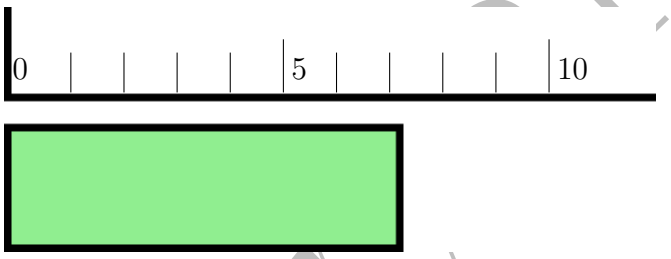


Figure 16.2: Different measurement?

the '7' and the '8' of the scale. The question is: *Can these two objects be distinguished?*

The first object is closer to the '8' than the '7', while the second object is closer to the '7' than the '8', so we might say they can *possibly* be distinguished. What scientists usually do in this situation is to **estimate** one more digit than they know for sure. So, for the first one, I might estimate it to be 7.7 units. (Someone else might estimate it to be 7.8 units, but that's fine.) The second one I estimate to be 7.2 units. (Someone else might estimate it to be 7.3 units, but that's also fine.) The *precision measure* of a scale like this would be said to be *one half of the smallest spacing on the scale.* Since this scale has spacings 1 unit apart, the precision measure would be 0.5 units.

Examples of analog instrument precision measure

In order to use instruments, you need to understand how to operate and read them. Here are instructions on how to use two important ones.

Micrometer Caliper A **micrometer caliper** is shown in Fig. 16.3, and a reading from a micrometer scale is shown in Fig. 16.4. As with other non-digital instruments, you must estimate one more digit than you know. The uncertainty is one half of the smallest division, as it would be for any non-digital instrument.[1]

Always use the ratchet to close a micrometer caliper; never use the thimble because that would allow you to apply enough force to damage the caliper.

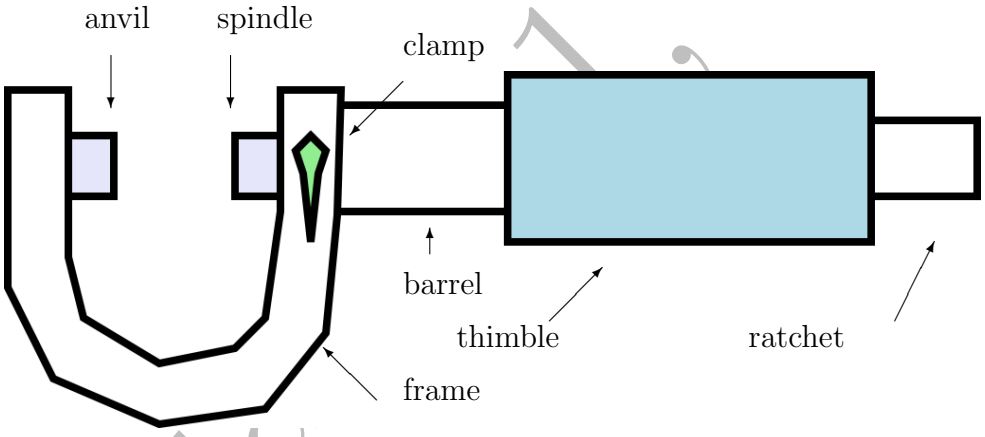


Figure 16.3: Micrometer Caliper

Figure 16.4 gives an example of a reading from a micrometer scale.

While the micrometer scale is linear, it is a little different than a regular scale because there are actually two distinct scales which must be read to determine the measurement. There is a horizontal scale on the barrel, which counts rotations of the thimble, and a vertical scale on the thimble, which gives the fractional part of the reading. (Usually the numbers on the thibmle will go up to 50, meaning that each complete rotation of the thimble represents a change of 0.5 of the units of the barrel scale.)

[1]The reason a micrometer is so named is that when the main scale is in millimetres, the digit estimated will be in thousandths of millimetres; ie. in "micrometres".

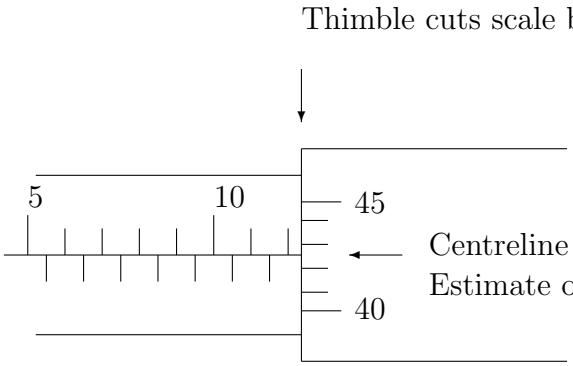


Figure 16.4: Reading a Micrometer Scale

Vernier Caliper A **Vernier caliper** is shown in Fig. 16.5. A reading from a vernier scale is shown in Fig. 16.6. The example shown is fairly simple. It is possible to have vernier scales which are more complicated, but the principle of operation is the same. The important thing to understand is the purpose of the two scales, and how to tell which is which.

Remember that the precision measure is the smallest difference between two measurements which can be distinguished, so that if the difference between two measurements is less than the precision measure they are the same. (Or, to be more correct, we say that they are *"the same within experimental uncertainty"*.

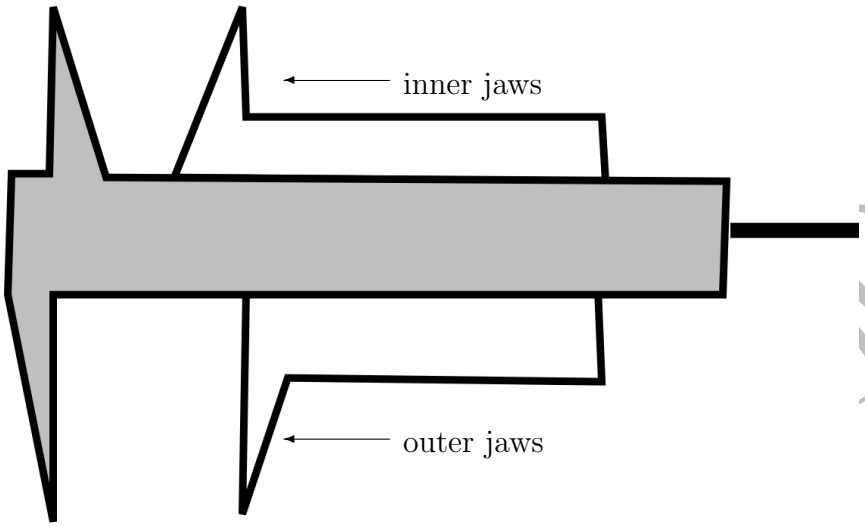
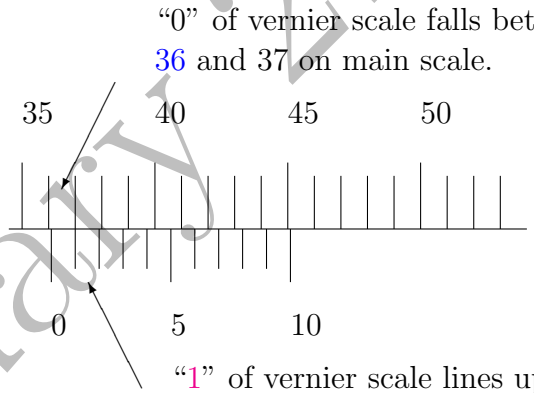


Figure 16.5: Vernier Caliper



The final reading is 36.1 ± 0.1 (of whatever appropriate units).

Figure 16.6: Reading a Vernier Scale

16.1.3 Precision and Measurement Uncertainty

Since the precision measure is the smallest difference between two measurements which can be distinguished, then we can say that a measurement made

with a given instrument will have an **uncertainty** equal to the precision measure. This means that another measurement would have to be either smaller or larger than this one by *at least* the precision measure to be distinguished from it. For this reason, the way that we usually record a measurement is

$$value \quad \pm \quad precision\ measure$$

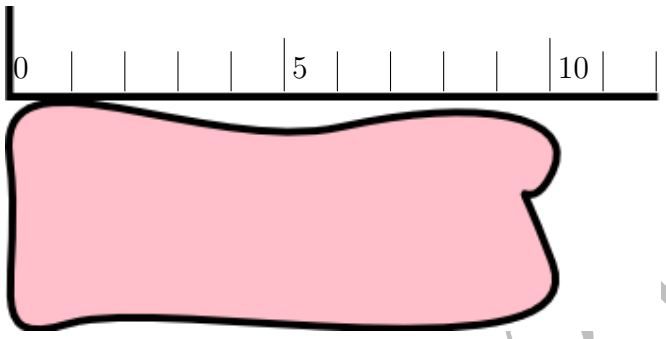


Figure 16.7: Measuring a realistic object

16.1.4 Effective Uncertainties

Reality is not always precise

Many times when we take measurements the precision measure of the instrument doesn't really matter, since we're trying to measure something "fuzzy". Look at the object in Figure 16.7. Since it doesn't have straight edges, measuring the length is a bit problematic. In fact, what we want to measure may depend on why we want to measure it. Consider these questions:

1. If we used this to prop open a window, how big a window opening could we have?

2. If we wanted to determine the area of the object, what length would we want?

3. If we wanted to use the edge for drawing straight lines on paper, what length of line could we draw?

In a situation where the precision measure isn't really the limitation on the precision of a measurement, we *estimate* a **realistic** or **effective** uncertainty based on whatever sort of limits make sense.

If we're going to give a realistic (or effective) uncertainty, it's going to have to be bigger than the precision measure, because the precision measure is the best we can do with the instrument. (Remember that if the difference between two measurements is less than the precision measure they are the same.)

16.1.5 Accuracy

The **accuracy** of an instrument refers to *how close a measurement is to the true value of the quantity being measured.* Usually if an instrument is inaccurate it is due to one of two factors:

- It doesn't read zero when it should.
- Readings that should be something other than zero are off by amounts that depend on the readings themselves.

The first of these is known as **zero error** and the second is known as *calibration error* or **linearity error**.

Zero error

Some measuring instruments have a certain **zero error** associated with them. This is particularly true of micrometer calipers, as used in this experiment. Take note of the *actual* reading of the instrument when the *expected* value would be zero. (For example, close the calipers and record the reading you get with its uncertainty.) Subsequent measurements should be corrected by adding or subtracting the zero error as appropriate. For some instruments, zero error will be easy to determine. For others, it may be very difficult.

Note that since the zero error is a reading taken from the instrument, it has an uncertainty equal to the precision measure, like any other measurement would.

Calibration Error (or linearity error)

A good example of a linearity error would be seen in a metal ruler. Since metal expands when heated, if the ruler was used at very high or very low temperatures the readings would be correspondingly high or low. (On the other hand, a wooden ruler may absorb humidity from the air and expand, or dry out and contract.)

While the precision measure of an instrument can usually be obtained by simply looking at the instrument, the accuracy of an instrument can only be determined by using it to measure known reference quantities.

16.2 Recap

By the end of this exercise, you should understand the following terms:

- range of an instrument
- precision (contrast with accuracy)
- precision measure
- accuracy (contrast with precision)
- measurement uncertainty
- zero error
- effective uncertainty
- range of values for a measurement

In addition, you should be able to take readings and express them with their uncertainties using:

- linear scales
- micrometer scales
- vernier scales
- digital displays

Chapter 17

Uncertain Results

17.1 The most important part of a lab

The "Discussion of Errors" (or Uncertainties) section of a lab report is where you outline the *reasonable limits* which you place on your results. If you have done a professional job of your research, you should be prepared to defend your results. In other words, you should expect anyone else to get results which agree with yours; if not, you suspect theirs. In this context, you want to discuss sources of error which you have reason to believe are significant.

17.1.1 Operations with Uncertainties

When numbers, some or all of which are approximate, are combined by addition, subtraction, multiplication, or division, the uncertainty in the results due to the uncertainties in the data is given by the *range of possible calculated values based on the range of possible data values.*

Remember: Since uncertainties are an indication of the imprecise nature of a quantity, uncertainties are usually only expressed to one decimal place. (In other words, it doesn't make sense to have an extremely *precise* measure of the *imprecision* in a value!)

For instance, if we have two numbers with uncertainties, such as $x = 2 \pm 1$ and $y = 32.0 + 0.2$, then what that means is that x can be as small as 1 or as big as 3, while y can be as small as 31.8 or as big as 32.2 so adding them can give a result $x + y$ which can be as small as 32.8 or as big as 35.2, so that

the uncertainty in the answer is the sum of the two uncertainties. If we call the uncertainties in x and y Δx and Δy, then we can illustrate as follows:

Adding

$$\begin{array}{rrcccccc} & x & \pm & \Delta x & = & 2 & \pm & 1 \\ + & y & \pm & \Delta y & = & 32.0 & \pm & 0.2 \\ & (x+y) & \pm & ? & = & 34.0 & \pm & 1.2 \\ \hline & & & & = & (x+y) & \pm & (\Delta x + \Delta y) \end{array}$$

Thus

$$\Delta(x+y) = \Delta x + \Delta y \tag{17.1}$$

Thus $x+y$ can be between 32.8 and 35.2, as above. (Note that we should actually express this result as 34 ± 1 to keep the correct number of significant figures.)

Remember: Uncertainties are usually only expressed to one decimal place, and quantities are written with the last digit being the uncertain one.

Subtracting

If we subtract two numbers, the same sort of thing happens.

$$\begin{array}{rrcccccc} & x & \pm & \Delta x & = & 45.3 & \pm & 0.4 \\ - & y & \pm & \Delta y & = & -18.7 & \pm & 0.3 \\ & (x-y) & \pm & ? & = & 26.6 & \pm & 0.7 \\ \hline & & & & = & (x-y) & \pm & (\Delta x + \Delta y) \end{array}$$

Thus

$$\Delta(x-y) = \Delta x + \Delta y \tag{17.2}$$

Note that we still *add* the uncertainties, even though we *subtract* the quantities.

Multiplying

Multiplication and division are a little different. If a block of wood is found to have a mass of 1.00 ± 0.03 kg and a volume of 0.020 ± 0.001 m^3, then the **nominal** value of the density is $\frac{1.00\text{kg}}{0.020\text{m}^3} = 50.0\text{kg/m}^3$ and the uncertainty in its density may be determined as follows:

The mass given above indicates the mass is known to be *greater than or equal to* 0.97 kg, while the volume is known to be *less than or equal to* 0.021 m^3. Thus, the *minimum* density of the block is given by $\frac{0.97\text{kg}}{0.021\text{m}^3} = 46.2\text{kg/m}^3$. Similarly, the mass is known to be *less than or equal to* 1.03 kg, while the volume is known to be *greater than or equal to* 0.019 m^3. Thus, the *maximum* density of the block is given by $\frac{1.03\text{kg}}{0.019\text{m}^3} = 54.2\text{kg/m}^3$.

Notice that the above calculations do not give a symmetric range of uncertainties about the nominal value. This complicates matters, but if uncertainties are small compared to the quantities involved, the range is approximately symmetric and may be estimated as follows:

$$\begin{array}{rcccc}
x \pm \Delta x & = & 1.23 \pm 0.01 & = & 1.23 \\
 & & & \pm & (0.01/1.23 \times 100\%) \\
\times \quad y \pm \Delta y & = & \times 7.1 \pm 0.2 & = & \times 7.1 \\
 & & & \pm & (0.2/7.1 \times 100\%) \\
(x \times y) \pm ? & = & 8.733 \pm ? & \approx & 8.733 \\
 & & & \pm & ((0.01/1.23 + 0.2/7.1) \times 100\%) \\
 & & & \approx & 8.733 \\
 & & & \pm & ((0.01/1.23 + 0.2/7.1) \times 8.733) \\
 & & & \approx & 8.733 \\
 & & & \pm & 0.317 \\
\hline
 & & & \approx & (x \times y) \\
 & & & \pm & \left(\frac{\Delta x}{x} + \frac{\Delta y}{y}\right)(x \times y)
\end{array}$$

Thus

$$\Delta(x \times y) \approx \left(\frac{\Delta x}{x} + \frac{\Delta y}{y}\right)(x \times y) \tag{17.3}$$

So rather than adding *absolute* uncertainties, we add *relative* or *percent* uncertainties. (To the correct number of significant figures, the above result would be

$$x \times y \approx 8.7 \pm 0.3$$

with one figure of uncertainty and the last digit of the result being the uncertain one.)

If you're a purist, or if the uncertainties are not small, then the uncertainty in the density can then be estimated in two obvious ways;

1. the *greater of the two* differences between the maximum and minimum and the accepted values

2. (or the maximum and minimum values can both be quoted, which is more precise, but can be cumbersome if subsequent calculations are necessary.)

(In the previous example, the first method would give an uncertainty of 4.2 kg/m^3.)

Dividing

Division is similar to multiplication, as subtraction was similar to addition.

$$\begin{array}{rcrcl}
x \pm \Delta x & = & 7.6 \pm 0.8 & = & 7.6 \\
 & & & \pm & (0.8/7.6 \times 100\%) \\
\div y \pm \Delta y & = & \div 2.5 \pm 0.1 & = & \div 2.5 \\
 & & & \pm & (0.1/2.5 \times 100\%) \\
(x \div y) \pm ? & = & 3.04 \pm ? & \approx & 3.04 \\
 & & & \pm & ((0.8/7.6 + 0.1/2.5) \times 100\%) \\
 & & & \approx & 3.04 \\
 & & & \pm & ((0.8/7.6 + 0.1/2.5) \times 3.04) \\
 & & & \approx & 3.04 \\
 & & & \pm & 0.4416 \\
\hline
 & & & = & (x \div y) \\
 & & & \pm & \left(\frac{\Delta x}{x} + \frac{\Delta y}{y}\right)(x \div y)
\end{array}$$

Thus

$$\Delta(x \div y) \approx \left(\frac{\Delta x}{x} + \frac{\Delta y}{y}\right)(x \div y) \tag{17.4}$$

(To the correct number of significant figures, the above result would be

$$x \div y \approx 3.0 \pm 0.4$$

with one figure of uncertainty and the last digit of the result being the uncertain one.)

Determining Uncertainties in Functions Algebraically

Consider a function as shown in Figure 17.1. If we want the know the uncertainty in $f(x)$ at a point x, what we mean is that we want to know *the difference between* $f(x + \Delta x)$ *and* $f(x)$.

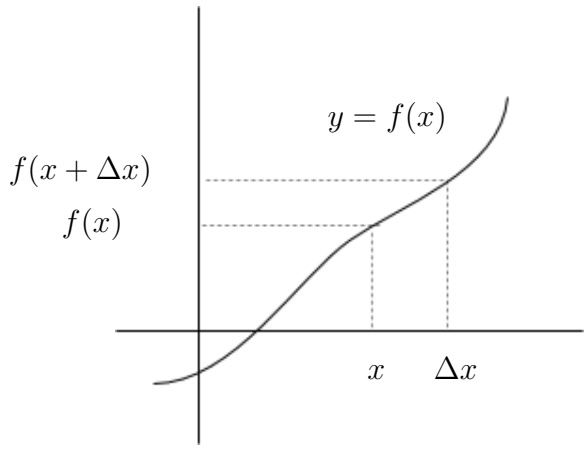


Figure 17.1: Uncertainty in a Function of x

If we take a closer look at the function, like in Figure 17.2, we can see that if Δx is small, then the difference between the function and its tangent line will be small. We can then say that

$$f(x) + f'(x) \times \Delta x \approx f(x + \Delta x)$$

or

$$\Delta f(x) \approx f'(x) \times \Delta x$$

For a function with a negative slope, the result would be similar, but the sign would change, so we write the rule with absolute value bars like this

$$\Delta f(x) \approx |f'(x)| \, \Delta x \tag{17.5}$$

to give an uncertainty which is positive.[1] Remember that uncertainties are usually rounded to one significant figure, so this approximation is generally valid.

[1]Now our use of the Δ symbol for uncertainties should make sense; in this example it has been used as in calculus to indicate "a small change in", but for experimental quantities, "small changes" are the result of uncertainties.

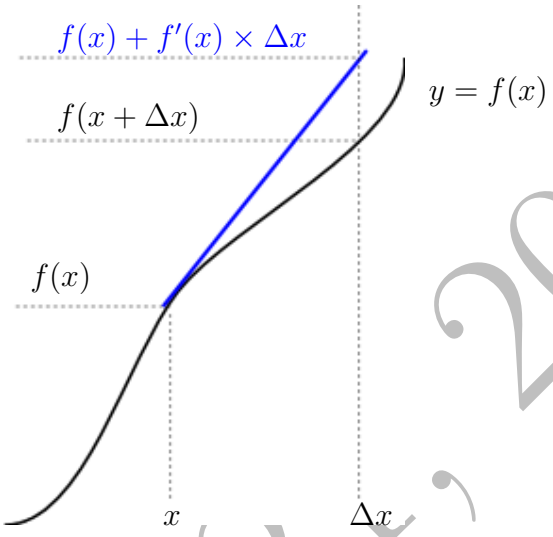


Figure 17.2: Closer View of Figure 17.1

Example: Marble volume Here is an example. Suppose we measure the diameter of a marble, d, with an uncertainty Δd, then quantities such as the volume derived from d will also have an uncertainty. Since

$$V = \frac{4}{3}\pi\left(\frac{d}{2}\right)^3$$

then

$$V' = 2\pi\left(\frac{d}{2}\right)^2 = \frac{\pi}{2}d^2$$

and so

$$\Delta V \approx \left|\frac{\pi}{2}d^2\right| \Delta d$$

If we have a value of $d = 1.0 \pm 0.1$ cm, then $\Delta V = 0.157$ cm^3 by this method. Rounded to one significant figure gives $\Delta V \approx 0.2$ cm^3.

Determining Uncertainties in Functions by Inspection

Note: In the following section and elsewhere in the manual, the notation Δx is used to mean *"the uncertainty in x"*.

When we have a measurement of $2.0 \pm 0.3\ cm$, this means that the *maximum* value it can have is $2.0 + 0.3cm$. The uncertainty is the difference between this maximum value and the *nominal* value (ie. the one with no uncertainty). We could also say that the *minimum* value it can have is $2.0 - 0.3\ cm$, and the uncertainty is the difference between the nominal value and this maximum value. Thus if we want to find the uncertainty in a function, $f(x)$, we can say that

$$\Delta f(x) \approx f_{max} - f \tag{17.6}$$

or

$$\Delta f(x) \approx f - f_{min} \tag{17.7}$$

where f_{max} is the same function with x replaced by *either* $x + \Delta x$ or $x - \Delta x$; whichever makes f bigger, and f_{min} is the same function with x replaced by *either* $x + \Delta x$ or $x - \Delta x$; whichever makes f *smaller*. (The approximately equals sign is to reflect the fact that these two values may not be quite the same, depending on the function f.) For instance, if

$$f(x) = x^2 + 5$$

then clearly, if x is positive, then replacing x by $x + \Delta x$ will make f a maximum.

$$f_{max} = f(x + \Delta x) = (x + \Delta x)^2 + 5$$

and so

$$\Delta f(x) \approx f_{max} - f = f(x + \Delta x) - f(x) = \left((x + \Delta x)^2 + 5\right) - \left(x^2 + 5\right)$$

On the other hand, if we wanted to find the uncertainty in

$$g(t) = \frac{1}{\sqrt{t}}$$

then, if t is positive, then replacing t by $t - \Delta t$ will make g a maximum.

$$g_{max} = g(t - \Delta t) = \frac{1}{\sqrt{(t - \Delta t)}}$$

and so

$$\Delta g(t) \approx g_{max} - g = g(t - \Delta t) - g(t) = \left(\frac{1}{\sqrt{(t - \Delta t)}}\right) - \left(\frac{1}{\sqrt{t}}\right)$$

If we had a function of two variables,

$$h(w, z) = \frac{\sqrt{w}}{z^2}$$

then we want to replace *each* quantity with the appropriate value in order to maximize the total, so if w and z are both positive,

$$h_{max} = \frac{\sqrt{(w + \Delta w)}}{(z - \Delta z)^2}$$

and thus

$$\Delta h \approx h_{max} - h = \frac{\sqrt{(w + \Delta w)}}{(z - \Delta z)^2} - \frac{\sqrt{w}}{z^2}$$

Notice in each of these cases, it was necessary to restrict the range of the variable in order to determine whether the uncertainty should be added or subtracted in order to maximize the result. In an experiment, usually your data will automatically be restricted in certain ways. (For instance, masses are always positive.)

Example:Marble volume Using the above example of the volume of a marble,

$$\Delta V \approx V(d + \Delta d) - V(d)$$

Since

$$V = \frac{4}{3}\pi\left(\frac{d}{2}\right)^3$$

then

$$\Delta V \approx \frac{4}{3}\pi\left(\frac{d + \Delta d}{2}\right)^3 - \frac{4}{3}\pi\left(\frac{d}{2}\right)^3$$

If we have a value of $d = 1.0 \pm 0.1$ cm, then $\Delta V = 0.173$ cm^3 by this method. Rounded to one significant figure gives $\Delta V \approx 0.2$ cm^3 as the value to be quoted.

Determining Uncertainties by Trial and Error

For a function $f(x, y)$, the uncertainty in f will be given by the *biggest* of

$$|f(x + \Delta x, y + \Delta y) - f(x, y)|$$

or

$$|f(x - \Delta x, y + \Delta y) - f(x, y)|$$

or

$$|f(x + \Delta x, y - \Delta y) - f(x, y)|$$

or

$$|f(x - \Delta x, y - \Delta y) - f(x, y)|$$

Note that for each variable with an uncertainty, the number of possibilities doubles. In most cases, common sense will tell you which one is going to be the important one, but things like the sign of numbers involved, etc. will matter a lot! For example, if you are adding two positive quantities, then the first or fourth above will give the same (correct) answer. However, if one quantity is negative, then the second and third will be correct.

The advantage of knowing this method is that it always works. Sometimes it may be easier to go through this approach than to do all of the algebra needed for a complicated expression, especially if common sense makes it easy to see which combination of signs gives the correct answer.

Determining Uncertainties Algebraically

To summarize, the uncertainty in results can *usually* be calculated as in the following examples (if the percentage uncertainties in the data are small):

$$\begin{aligned}
(a) \qquad & \Delta(A + B) = (\Delta A + \Delta B) \\
(b) \qquad & \Delta(A - B) = (\Delta A + \Delta B) \\
(c) \qquad & \Delta(A \times B) \approx |AB| \left(\left|\tfrac{\Delta A}{A}\right| + \left|\tfrac{\Delta B}{B}\right| \right) \\
(d) \qquad & \Delta(\tfrac{A}{B}) \approx \left|\tfrac{A}{B}\right| \left(\left|\tfrac{\Delta A}{A}\right| + \left|\tfrac{\Delta B}{B}\right| \right) \\
(e) \qquad & \Delta f(A \pm \Delta A) \approx |f'(A)|\, \Delta A
\end{aligned}$$

Note that the first two rules above *always* hold true.

To put it another way, when *adding* or *subtracting*, you *add absolute* uncertainties. When *multiplying* or *dividing*, you *add percent* or *relative* uncertainties. Note that for the last rule above that angles and their uncertainties must be in *radians* for the differentiation to be correct! (In the examples above, absolute value signs were omitted since all positive quantities were used.) (Some specific uncertainty results can be found in *Appendix I*.)

Remember that a quantity and its uncertainty should always have the same units, so you can check units when calculating uncertainties to avoid mistakes.

Two important corrollaries: constants and powers The above rules can be used to derive the results for two very common situations;

- multiplying a quantity with an uncertainty by a constant
- raising a quantity with an uncertainty to a power

In the first case, a constant can be thought of as a number with *no* uncertainty. The product rule above is

$$\Delta(A \times B) \approx |AB| \left(\left| \frac{\Delta A}{A} \right| + \left| \frac{\Delta B}{B} \right| \right)$$

If A is a constant, then $\Delta A = 0$, so

$$\Delta(A \times B) \approx |AB| \left(\left| \frac{\cancel{\Delta A}}{\cancel{A}} \right| + \left| \frac{\Delta B}{B} \right| \right) = \left| A\cancel{B} \right| \left(\left| \frac{\Delta B}{\cancel{B}} \right| \right) = |A \Delta B| = |A| \, \Delta B$$

In the second case, the product rule is:

$$\Delta f(A \pm \Delta A) \approx |f'(A)| \, \Delta A$$

and so if

$$f(A) = A^n$$

then

$$f'(A) = nA^{n-1}$$

and so

$$\Delta(A \pm \Delta A)^n \approx \left| nA^{n-1} \right| \Delta A$$

Example: Marble volume If we have a value of $d = 1.0 \pm 0.1$ cm, as used previously, then $\Delta V = 0.157$ cm^3 by this method.

Mathematically, this result and the previous one are equal if $\Delta d << d$. You can derive this using the **binomial approximation**, which simply means multiplying it out and discarding and terms with two or more Δ terms multiplied together; for instance $\Delta A \Delta B \approx 0$

Choosing Algebra or Inspection

Since uncertainties are usually only expressed to one decimal place, then small differences given by different methods of calculation, (ie. inspection or algebra), do not matter.

Example: Marble volume Using the previous example of the marble, if we have a value of $d = 1.0 \pm 0.1$ cm, then $\Delta V = 0.173$ cm^3 by the inspection method. Rounded to one significant figure gives $\Delta V \approx 0.2$ cm^3 as the value to be quoted. By the algebraic method, $\Delta V = 0.157$ cm^3. Rounded to one significant figure gives $\Delta V \approx 0.2$ cm^3, which is the same as that given by the previous method. So in this example a 10% uncertainty in d was still small enough to give the same result (to one significant figure) by both methods.

Sensitivity of Total Uncertainty to Individual Uncertainties

When you discuss sources of uncertainty in an experiment, it is important to recognize which ones contributed most to the uncertainty in the final result. In order to determine this, proceed as follows:

1. Write out the equation for the uncertainty in the result, using whichever method you prefer.

2. For each of the quantities in the equation which have an uncertainty, calculate the uncertainty in the result *which you get if all of the other uncertainties are zero.*

3. Arrange the quantities in descending order based on the size of the uncertainties calculated. The higher in the list a quantity is, the greater it's contribution to the total uncertainty.

The sizes of these uncertainties should tell you which factors need to be considered, remembering that only quantities contributing 10% or more to the total uncertainty matter. For example, from before we had a function of two variables,

$$h(w,z) = \frac{\sqrt{w}}{z^2}$$

so by inspection, its uncertainty is given by

$$\Delta h \approx h_{max} - h = \frac{\sqrt{(w+\Delta w)}}{(z-\Delta z)^2} - \frac{\sqrt{w}}{z^2}$$

So we can compute

$$\Delta h_w \approx \frac{\sqrt{(w+\Delta w)}}{z^2} - \frac{\sqrt{w}}{z^2}$$

and

$$\Delta h_z \approx \frac{\sqrt{w}}{(z-\Delta z)^2} - \frac{\sqrt{w}}{z^2}$$

Note that in the first equation, all of the Δz terms are gone, and in the second, all of the Δw terms are gone. By the algebraic method,

$$\Delta h \approx \frac{\sqrt{w}}{z^2}\left(\frac{\Delta w}{2w} + \frac{2\Delta z}{z}\right)$$

and so

$$\Delta h_w \approx \frac{\sqrt{w}}{z^2}\left(\frac{\Delta w}{2w}\right)$$

and

$$\Delta h_z \approx \frac{\sqrt{w}}{z^2}\left(\frac{2\Delta z}{z}\right)$$

Note that *until you plug values into these equations, you can't tell which uncertainty contribution is larger.*

In the above example, if we use values of $w = 1.00 \pm 0.01$ and $z = 2.00 \pm 0.02$, then the proportional uncertainties in both w and z are the same, 1%. However, using either inspection or the algebraic method, $\Delta h = 0.006$, and $\Delta h_w = 0.001$ while $\Delta h_z = 0.005$; in other words, the uncertainty in the *result* due to Δz is five times the uncertainty due to Δw! (As you get more used to uncertainty calculations, you should realize this is because z is raised to a higher power than w, and so its uncertainty counts for more.) In order to improve this experiment, it would be more important to try and reduce Δz than it would be to try and reduce Δw.

Simplifying Uncertainties

Uncertainty calculations can get quite involved if there are several quantities involved. However, since uncertainties are usually only carried to one or two significant figures at most, there is little value in carrying uncertainties through calculations if they do not contribute significantly to the total.

> You do not need to carry uncertainties through if they do not contribute more than 10% of the total uncertainty, since uncertainties are usually only expressed to one decimal place. (However, be sure to give bounds for these uncertainties when you do this.)

Note that this shows a difference between doing calculations by hand versus using a spreadsheet. If you are doing calculations by hand, it makes sense to drop insignificant uncertainties like this.

> If you're using a spreadsheet in order to allow you to change the data and recalculate, it may be worth carrying all uncertainties through in case some of them may be more significant for different data.

17.1.2 Uncertainties and Final Results

When an experiment is performed, it is crucial to determine whether or not the results *make sense*. In other words, do any calculated quantities fall within a "reasonable" range?

The reason for doing calculations with uncertainties is so that uncertainties in *final answers* can be obtained. If, for instance, a physical constant was measured, the calculated uncertainty determines the range around the calculated value in which one would expect to find the "theoretical" value. *If* the theoretical value falls within this range, then we say that our results *agree* with the theory *within our experimental uncertainty*.

For instance, if we perform an experiment and get a value for the acceleration due to gravity of $g = 9.5 \pm 0.5\text{m/s}^2$ then we can say that we say that our values agrees with the accepted value of $g = 9.8\text{m/s}^2$ *within our experimental uncertainty*.

If we have two values to compare, such as initial and final momentum to determine whether momentum was conserved, then we see if the ranges given

by the two uncertainties overlap. In other words, if there is a value or range of values common to both, then they agree within experimental uncertainty.

So if an experiment gives us a value of $p_i = 51.2 \pm 0.7$ kg-m/s and $p_f = 50.8 \pm 0.5$ kg-m/s, then we would say the values agree within experimental uncertainty since the range from 50.5 kg-m/s $\rightarrow$ 51.3 kg-m/s is common to both. Since what we were studying was the conservation of momentum, then we would say that in this case momentum was conserved within experimental uncertainty. Note that if both uncertainties were 0.1 kg-m/s, then our results would *not* agree and we would say that momentum was *not* conserved within experimental uncertainty.

Mathematically, if two quantities a and b, with uncertainties Δa and Δb are compared, they can be considered to agree within their uncertainties if

$$|a - b| \leq \Delta a + \Delta b \tag{17.8}$$

A constant given with no uncertainty given can usually be assumed to have an uncertainty of zero.

If we need to compare 3 or more values this becomes more complex.

If two quantities agree within experimental error, this means that the discrepancy between experiment and theory can be readily accounted for on the basis of measurement uncertainties which are known. If the theoretical value does not fall within this range, then we say that our results *do not agree* with the theory within experimental uncertainty. In this situation, we cannot account for the discrepancy on the basis of measurement uncertainties alone, and so some other factors must be responsible.

If two numbers do not agree within experimental error, then the *percentage difference* between the experimental and theoretical values must be calculated as follows:

$$Percent\ Difference = \left| \frac{theoretical - experimental}{theoretical} \right| \times 100\% \tag{17.9}$$

Remember: Only calculate the percent difference if your results do <u>not</u> agree within experimental error.

In our example above, we would *not* calculate the percentage difference between our calculated value for the acceleration due to gravity of $g = 9.5 \pm 0.5\text{m/s}^2$ and the accepted value of $g = 9.8\text{m/s}^2$ since they agree within our experimental uncertainty.

Often instead of comparing an experimental value to a theoretical one, we are asked to test a law such as the Conservation of Energy. In this case, what we must do is to compare the initial and final energies of the system in the manner just outlined.[2] If the values agree, then we can say that energy was conserved, and if the values don't agree then it wasn't. In that case we would calculate the percentage difference as follows:

$$Percent\ Difference = \left| \frac{initial - final}{initial} \right| \times 100\% \qquad (17.10)$$

Significant Figures in Final Results

Always express final answers with absolute uncertainties rather than percent uncertainties. Also, always quote final answers with one significant digit of uncertainty, and round the answers so that the least significant digit quoted is the uncertain one. This follows the same rule for significant figures in measured values.

Even though you want to round off your final answers to the right number of decimal places, don't round off in the middle of calculations since this will introduce errors of its own.

17.1.3 Discussion of Uncertainties

In an experiment, with each quantity measured, it is necessary to consider all of the possible sources of error *in that quantity*, so that a realistic uncertainty can be stated for that measurement. The "Discussion of Uncertainties" (or "Discussion of Errors") is the section of the lab report where this process can be explained.

[2] There is another possibility which you may consider. Suppose you compare the *change in energy* to its expected value of zero. In that case, *any* non-zero change would result in infinite percent difference, which is mathematically correct but not terribly meaningful physically.

Discussions of sources of error should always be made as concrete as possible. That means they should be use specific numerical values and relate to specific experimental quantities. For instance, if you are going to speak about possible air currents affecting the path of the ball in a fre–fall experiment, you must reduce it to a finite change in either the fall *time* or the *height*.

Relative Size of Uncertainties

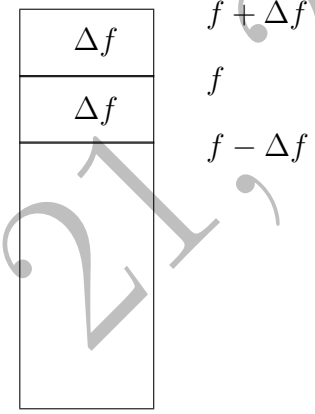


Figure 17.3: Relative Size of Quantity and its Uncertainty

The uncertainties which matter most in an experiment are those which contribute most to the uncertainty in the final result. Consider Figure 17.4, which may be seen as a magnification of one of the bands in Figure 17.3.

Figure 17.4: Contributions of Various Sources to Total Uncertainty

If the big rectangle represents the uncertainty in the final result, and the smaller rectangles inside represent contributions to the total from various sources, then one source contributes almost half of the total uncertainty in the result. The first two sources contribute about 75% of the total, so that all of the other sources *combined* only contribute about 25%. If we want to improve the experiment, we should try to address the factors contributing most. Similarly, in discussing our uncertainties, the biggest ones deserve most attention. In fact, since uncertainties are rounded to one decimal place, any uncertainty contributing less than 10% to the final uncertainty is basically

irrelevant. The only reason to discuss such uncertainties is to explain why they are not significant.

Types of Errors

There are 3 major "categories" of sources of errors, in order of importance;

1. **Measurable uncertainties**-these are usually the biggest. The precision measure of each instrument used must always be recorded with every measurement. If "pre-measured" quantities are used, (such as standard masses), then there will usually be uncertainties given for these as well. If physical constants are given they may have uncertainties given for them, (such as the variation in the acceleration due to gravity by height above sea level, latitude, etc.) Where the realistic uncertainty in a quantity comes from any of these, (which will often be the case), you do not usually need to refer to them in your discussion. However, if there are any which contribute *greatly* to the uncertainty in your results, you should discuss them. For example, when you measure the mass of an object with a balance, then if the precision measure is the uncertainty used in your calculations, you don't need to discuss it, *unless* it is one of the *biggest* uncertainties in your calculations. *Keep in mind that without these values being given, it is impossible to tell whether any of the following sources of error are significant or not.*

2. **Bounded uncertainties**-these are things *which you observed*, and *have put limits on* and *usually* are much smaller than those in the group above. (Remember that since uncertainties are ultimately rounded to one significant digit, any which contribute less than 10% to the total uncertainty can be ignored.) Since you have observed them, you can give some estimate of how much effect they may have. For instance, suppose you measure the length of a table with a metre stick, and notice that the ends of the table are not exactly smooth and straight. If you can find a way to measure the variation in the length of the table due to this, then you can incorporate this into your uncertainty (if it is big enough) and discuss it.

A *plausible* error is one which can be tested. If you cannot figure out how to test for an error, it is not worth discussing. (Putting a bound on an error implies some method of testing for its existence, even if you are not able to do it at the present time.)

3. **Blatant filler**-these are things you may be tempted to throw in to sound more impressive. *Don't!!! If you did not observe them, don't discuss them.* If you suggest the gravitational pull of Jupiter is affecting your results, you'd better be prepared to show evidence (such as getting consistently different results at different times of day as the Earth rotates and so changes the angle of Jupiter's pull.) Do you even *know* in which direction the pull of Jupiter would be???

If you are going to discuss a source of uncertainty, then you must either have included it in your calculations, or given some reasonable bounds on its size. If you haven't done either of those, forget it!

You must discuss at least one source of systematic error in your report, even if you reject it as insignificant, in order to indicate how it would affect the results.

Reducing Errors

Whenever errors are discussed, you should suggest how they may be reduced or eliminated. There is a "hierarchy" of improvements which should be evident in your discussion. The following list starts with the best ideas, and progresses to less useful ones.

1. Be smart in the first place. You should never suggest you may have done something wrong in the lab; a professional who recognizes a mistake goes back and fixes it before producing a report. If you find yourself making a mistake which would seem likely to be repeated by other people, you may want to mention it in your report so that instructions may be clarified for the future.

2. Repeat the measurements once or twice *to check for consistency.* Repetition is a very *good* thing to do if your data are inconsistent or scattered. If certain values appear to be *incorrect*, you may want to repeat

them to make sure. If this seems to be true, and you feel a measurement was wrong, you should still include it in your report but explain why it was not used in your calculations. (This is probably similar to the previous one; *if* you think your data may be messed up, you should try to repeat it *before* you write your report, so this is not something you should be *suggesting* in your own report, although you should explain that you did it if you felt it was necessary.)

3. Change technique. It may be that a different way of doing things, using the *same equipment*, could (potentially) improve your results. If so, this should be explained.

   One example of this which may sound odd at first is to try and increase *the error and see what change is produced. For instance, if you neglected the mass of something in an experiment, you could increase that mass and then repeat the experiment. If the results do not change, then it is unlikely that the original mass had a significant effect.*

   Question: How big a change in the quantity in question (such as the mass just mentioned) should you try? Explain.

4. Make more *types* of observations. In some cases, monitoring certain things during the experiment may ensure they do not affect the results. This may be relevant in the case of "bounded uncertainties" above. It should be possible with equipment available in the lab. (For instance, if you are measuring the speed of sound, and the expected value is given at 25° C, then you might explain a discrepancy by the temperature being different. However, in this case, if you think the temperature may have affected your results, then you should check a thermometer to get the actual temperature during the experiment to suggest whether or not that was likely to have caused an effect.)

5. Repeat the measurements *to average the results.* While it is always good to repeat measurements, there is a law of diminishing returns. (In other words, repeating measurements a few times will give you a lot of information about how consistent your results are; repeating them many *more* times will not tell you as much. That is why the standard deviation of the mean decreases as $1/\sqrt{n}$, where n is the number of measurements; as n gets bigger, the change happens more slowly.) In fact, depending on the uncertainties involved, repetition at some point

is of no value. (That is when the standard deviation of the mean gets smaller than the uncertainty in the individual measurements. At that point you cannot improve without using a more precise instrument, no matter how many times you repeat the experiment.)

6. Change equipment; this is a *last resort*. Since this in essence means doing a different experiment, it is least desirable, and least relevant. Your goal is to produce the best results possible *with the equipment available*.

Ridiculous Errors

Certain errors crop up from time to time in peoples' reports without any justification. The point of your discussion is to *support* your results, placing *reasonable* bounds on them, not to absolve yourself of responsibility for them. Would you want to hire people who did not have faith in their own research? Including errors merely to "pad" your report is not good; one realistic source of error with justification is better than a page full of meaningless ones. Following are some commonly occurring meaningless ones.

- *"..human error..."*

  This is the *most irritating* statement you can make; you should have read over the instructions beforehand until you knew what was required, and then performed the experiment to the best of your ability. If you didn't you were being unprofessional and are wasting the reader's time. After doing your calculations, you should be able to tell from your results if they make sense. If not, you should go back and correct your errors. (Note something like reaction time does *not* fall into this category, because it is well-defined and can easily be measured. Vague, undefined errors are the big no-no.)

- *"..parallax..."*

  Parallax is the error you get from looking at a scale like a speedometer or a clock from the side; the position of the hands will appear different depending on your angle. With just about any scale I've seen, I'd be hard pressed to get an error of more than $5 \rightarrow 10\%$ from parallax (and the latter very rarely). Even that would only occur if I was deliberately *trying* to observe off-axis. Unless there is some reason that you cannot eliminate it, don't ascribe any significant error to it.

- *"..component values may not have been as stated..."*

  Usually people say this about masses, etc. I'm tempted to say *"Well, DUH! "* but I won't. *Of course* if given values are incorrect then calculations will be in error, but unless you have *evidence* for a specific value being wrong, (which should include some bounds on *how* wrong it could be), then it is just wild speculation. (You may allow *reasonable* uncertainties for these given values if you justify them.) Of course, suggesting equipment was damaged or broken is in this same category. If you understand what is going on, you should be able to tell if the equipment is functioning correctly. If it isn't, you should fix it or replace it (unless it's not working because you are not using it correctly; in that case, see "human error" above.) If it's possible you have broken it, you should bring this to the attention of the lab demonstrator, and be *very sure* you know how to use it properly before trying again with new equipment.

A Note on Human Errors

By now you are probably wondering why human error is so bad, even though humans have to make judgments in experiments, which will certainly contribute to uncertainties in the results. The problem is vague unspecified "human error" which is more of a disclaimer than a real thoughtful explanation. *If* you had to judge the time when an object stopped moving, for instance, you *can* discuss the judgment required, but in that case you should be able to determine concrete bounds for the uncertainties introduced, rather than suggesting some vague idea that your results may be meaningless.

A rule of thumb to follow in deciding whether a particular type of "human error" is valid is this; if it is something which *you* may have done wrong, that is not valid. If it is a limitation which *anyone* would have doing the experiment, then it is OK, provided you bound it. (But don't call it "human error"; be specific about what judgment is involved.)

Chapter 18

Graphical Data Analysis

18.1 Theory

18.1.1 Correlation

A **correlation** exists if a graph of two variables shows some sort of trend; in other words, knowing the value of one variable gives some prediction about the value of the other variable. In this case, the easiest way to view the comparison is by an **x-y graph**, where one variable is plotted on each axis.[1]

Figure 18.1 shows a graph where the correlation is obvious. All of the points fit on a straight line. This is rarely the case. More often the data look like Figure 18.2. This graph does not show as clear a correlation as the previous one, but the trend is still clear; points go from bottom left to top right. Knowing the value on one axis will allow a prediction of the value on the other axis with reasonable accuracy.

In Figure 18.3 there is no clear relationship between values on the two axes; knowing the value on one will give no estimate about the value on the other. This shows that the two variables are unrelated.

[1] Correlations don't have to be linear, but it's easier to illustrate them that way. Similarly, correlations can be negative (i.e. having a negative slope), but examples will usually be positive.

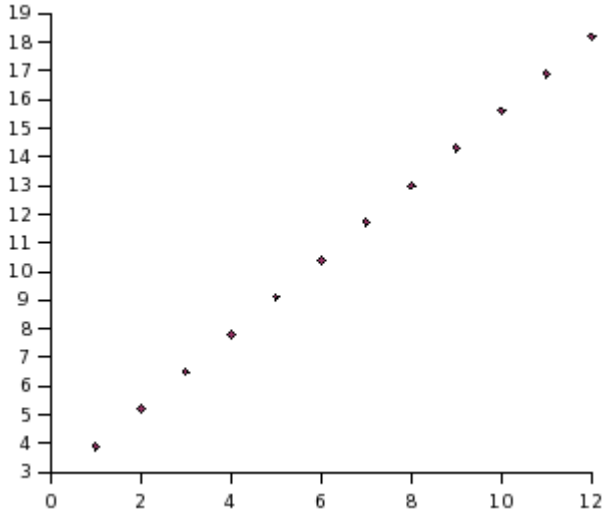


Figure 18.1: Obvious positive correlation

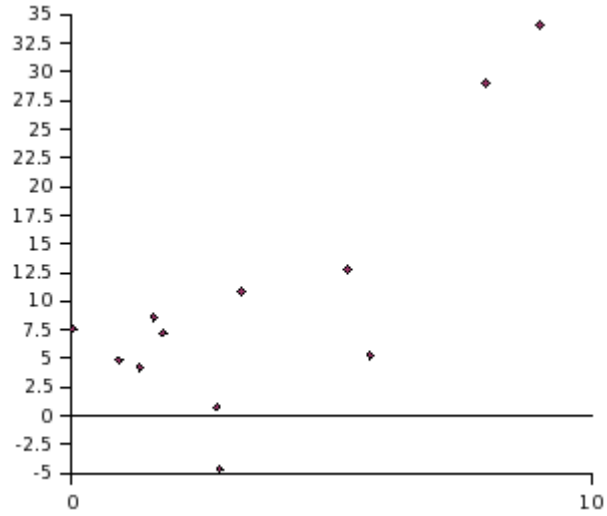


Figure 18.2: "Noisy" positive correlation

Correlation is not causation!

This is a point which cannot be overstated. The existence of a correlation doesn't indicate *why* the correlation exists. Here are a few examples:

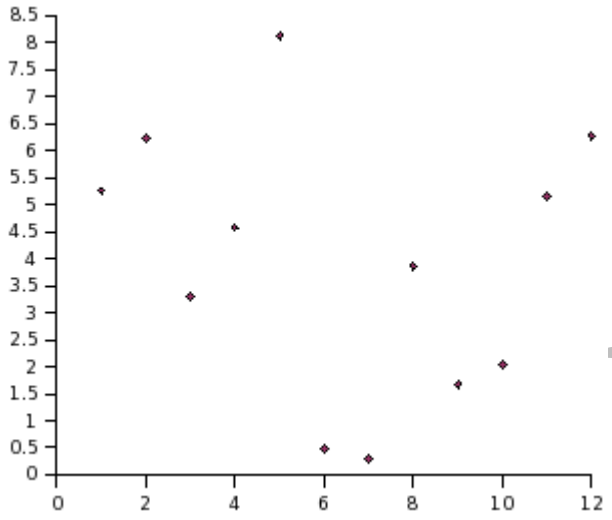


Figure 18.3: No correlation; random

1. Plotting number of degrees obtained versus income would show a positive correlation. This does *not* mean that rising income results in degrees being issued. The relationship is the other way around.[2]

2. A plot of vocabulary versus height would show a positive correlation. This does not mean that taller adults have bigger vocabularies than shorter adults; *children* are shorter than adults and have a smaller vocabulary. *Age* is the variable which is responsible for the changes in both height and vocabulary.

If the slope is zero within experimental uncertainty, then there is no correlation.

Correlation and prediction

One of the main reasons for looking for correlations in data, if not the *main* reason, is to be able to **predict** the value of one variable from the known value of the other. A strong correlation between two variables means one may be

[2]Of course, honorary degrees are occasionally given to very successful, (specifically, wealthy), people, although the number of these degrees is quite small.

used to predict the other. It doesn't matter whether the correlation is positive or negative. The weaker the correlation, the less precise the prediction can be. If there is no correlation, then by definition, one variable cannot be used to predict the other.

A strong correlation, either positive or negative, means one variable is a good predictor of the other.

18.1.2 Comparison

Sometimes, instead of having two numerical variables, you have one numerical variable for more than one data set, where the data sets are differentiated by some non-numerical parameter. For instance, in an election poll, you might show percentage of voters favouring each party, (a numerical variable), broken down by party, (a non-numerical parameter). In this case, probably the easiest way to view the comparison is by a **bar graph**, where each bar is for a different value of the non-numeric parameter. [3] Such a graph is shown in Figure 18.1. To turn this into a stacked bar graph, we need to modify

	Party			
Support	Liberal	Conservative	NDP	Green
Percent	32	25	10	4
Δ%	5	4	2	1

Table 18.1: Mythical Poll Results

the data slightly, by making a third row, which is a duplicate of the second, and changing the items in the first row by subtracting the values from the second row. This is shown in Figure 18.2. In this way, the minimum value for each row, (i.e. the nominal value minus the uncertainty), is given by the top of the lowest bar, the nominal value is given by the top of the middle bar, and the maximum value, (i.e. the nominal value plus the uncertainty), is given by the top of the top bar. The modified data can then be plotted to show how the different quantities compare. From Figure 18.4 we can see that support for Liberals and Conservatives is the same, within experimental

[3] If the numerical values are percentages, a pie chart may be more useful.

Support	Party			
	Liberal	Conservative	NDP	Green
Percent - Δ%	27	21	8	3
Δ%	5	4	2	1
Δ%	5	4	2	1

Table 18.2: Modified Data

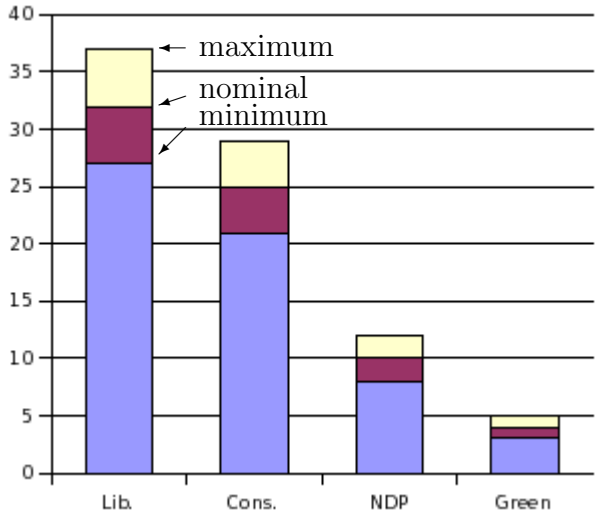


Figure 18.4: Comparing items

uncertainties, since a horizontal line can be drawn which passes through the uncertainty range for both. For any other combination, this is not the case.

If the values agree within experimental uncertainty, then they are the same.

18.2 Recap

By the end of this exercise, you should understand the following terms:

- correlation
- comparison

In addition, you should know what kind of graph to use to illustrate each.

Chapter 19

Graphs and Graphical Analysis

19.1 Introduction

One of the purposes of a scientific report is to present numerical information, ie. data and calculated results, in concise and meaningful ways. As with other parts of the report, the goal is to make the report as self-explanatory as possible. Ideally a person unfamiliar with the experiment should be able to understand the report without having to read the lab manual. (In your case, the reader can be assumed to be familiar with the general procedure of the experiment, but should not be expected to be intimately familiar with experiment-specific symbols. For instance, if you must measure the diameter of an object in the lab, and use the symbol d for it, be sure to state what d represents the first time it is used.)

A physical law is a mathematical relationship between measurable quantities, as has been stated earlier. A graph is a visual representation of such a relationship. In other words, a graph is always a representation of a particular mathematical relationship between the variables on the two axes; usually these relationships are made to be functions.

As a representation of how data are related, a graph will usually contain both data points and a **fitted curve** showing the function which the data should follow. (The term "curve" may include a straight line. In fact, it is often easiest to interpret results when an equation has been **linearized** so that the graph should be a straight line. Linearization will be discussed in Chapter 20, *"Linearizing Equations"* .)

With single values which are measured or calculated, when there is an

"expected" value, then uncertainties are used to determine how well the experimental value matches the expected value. For a set of data which should fit an equation, it is necessary to see how all points match the function. This is done using **error bars**, which will be discussed later. In essence, error bars allow one to observe how well each data point fits the curve or line on the graph. When parameters of an equation, such as the slope and y-intercept of a straight line, are determined from the data, (as will be discussed later), then those parameters will have uncertainties which represent the range of values needed to make all of the data points fit the curve.

19.2 Graphing

19.2.1 Data Tables

Often the data which is collected in an experiment is in a different form than that which must be plotted on a graph. (For instance, masses are measured but a graph requires weights.) In this case, the data which is to be plotted should be in a data table of its own. This is to make it easy for a reader to compare each point in the data table with its corresponding point on the graph. The data table should include the size of error bars for each point, in each dimension. Units in the table should be the same as on the graph.

Any graph must be plotted from data, which should be presented in tables. Tables should

- have ruled lines outside and separating columns, etc. to make it neat and easy to read
- have meaningful title and column headings
- not be split up by page breaks (ie. unless a table is bigger than a single page, it should all fit on one page.)
- have a number associated with it (such as *"Table 1"*) for reference elsewhere in the report, and a name, (such as *"Steel Ball Rolling down Incline"*) which makes it self–explanatory
- include the information required for any numerical data, ie. units, uncertainties, etc.

A sample is shown in Table 19.1.

i	x_i (cm)	Δx_i (cm)	t_i (s)	Δt_i (s)
1	0.40	0.03	0.0	0.1
2	0.77	0.04	2.0	0.1
3	1.35	0.04	2.7	0.1

Table 19.1: Position versus Time for Cart

19.2.2 Parts of a Graph

1. Title

   The title of a graph should make the graph somewhat self–explanatory aside from the lab. Something like "y vs. x" may be correct but redundant and useless if the person viewing the graph can read. "Object in Free Fall" would be more helpful as the reader may be able to figure out the *significance* of the graph herself.

2. Axis Labels

   As above, "m" and "l" are not as useful as "added mass (m) in grams", and "length of spring (l) in cm". In this case the words are meaningful, while the symbols are still shown to make it easy to find them in equations. Units *must* be included with axis labels.

3. Axis Scales

   The following 3 points are pertinent if you are plotting graphs "by hand". If you use a spreadsheet, these things are usually taken care of automatically.

   (a) Always choose the scales of the axes so that the data points will be spread out over as much of the plotting area as possible.

   (b) Choose the scales in a convenient manner. Scales that are easy to work with are to be preferred over scales such as ones where every small division corresponds to 0.3967 volts, for example. A better choice in such a case would be either 0.25, 0.50, or perhaps even 0.4 volts per division, the decision of which would be determined by the previous constraint. If you have *discrete*, i.e. integer, values on one axis, do not use scientific notation to represent those values.

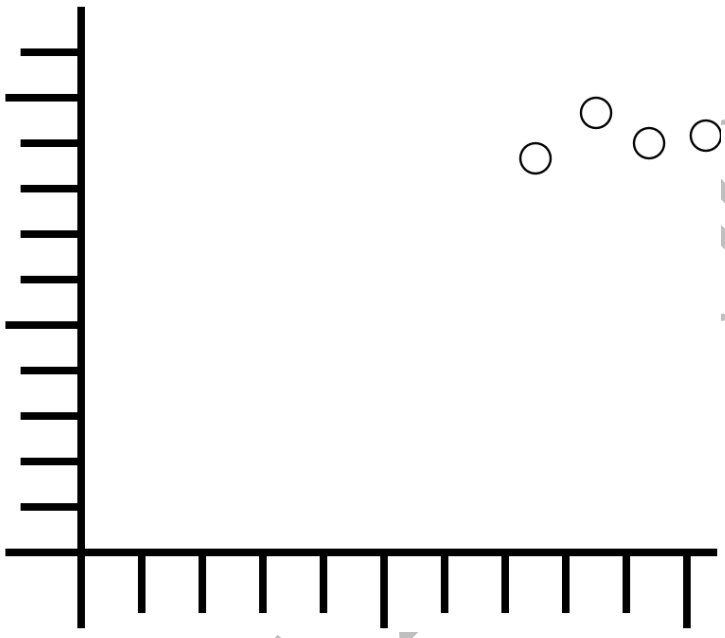


Figure 19.1: Wrong: Data, (not empty space), should fill most of the graph

(c) (0,0) does *not* have to be on graph – *data* should cover more than 1/4 of the graph area; if you need to extrapolate, do it numerically.

4. Plotting Points

Often, results obtained from graphs are slightly suspicious due to the simple fact that the experimenter has incorrectly plotted data points. If plotting by hand, be careful about this. Data points *must be* fitted with *error bars* to show uncertainties present in the data values. If the uncertainties in either or both dimensions are too small to show up on a particular graph, a note to that effect should be made on the graph so that the reader is aware of that fact.

Do not connect the points like a dot-to-dot drawing!

5. Points for Slope

   In calculating parameters from a graph, such as the slope, points on a line should be chosen which are *not* data points, *even if data points appear to fall directly on the line*; failure to follow this rule makes the actual line drawn irrelevant and misleading.

   When plotting points for the slope, a different symbol should be used from that used for data points to avoid confusion. The co–ordinates of these data points should be shown near the point as well for the reader's information. If one uses graph paper with a small enough grid, it may be possible to choose points for the slope which fall on the intersection of grid lines which simplifies the process of determining their co–ordinates. Of course, points for the slope should always be chosen as far apart as possible to minimize errors in calculation.

6. Error Bars

   Data points *must be* fitted with *error bars* to show uncertainties present in the data values. If the uncertainties in either or both dimensions are too small to show up on a particular graph, a note to that effect should be made on the graph so that the reader is aware of that fact. Uncertainties in quantities plotted on a graph are shown by error bars. Figure 19.2 shows a point with its error bars. The range of possible values for the data point in question actually includes any point bounded by the rectangle whose edges fall on the error bars. The size of the error bars is given by the uncertainties in both coordinates. (Actually, the point's true value is most likely to fall within the ellipse whose extents fall on the error bars. This is because it is unlikely that the x and y measurements are both in error by the maximum amount at the same time.) In fact, error bars may be in one or both directions, and they may even be different in the positive and negative directions.

Is the origin a data point?

Sometimes an experiment produces a graph which is expected to go through $(0, 0)$. In this case, whether to include the origin as a data point or not arises. There is a basic rule: *Include* $(0, 0)$ *as a data point only if you have measured it (like any other data point).* Often a graph which is *expected* to go through the origin will not do so due to some experimental factor which

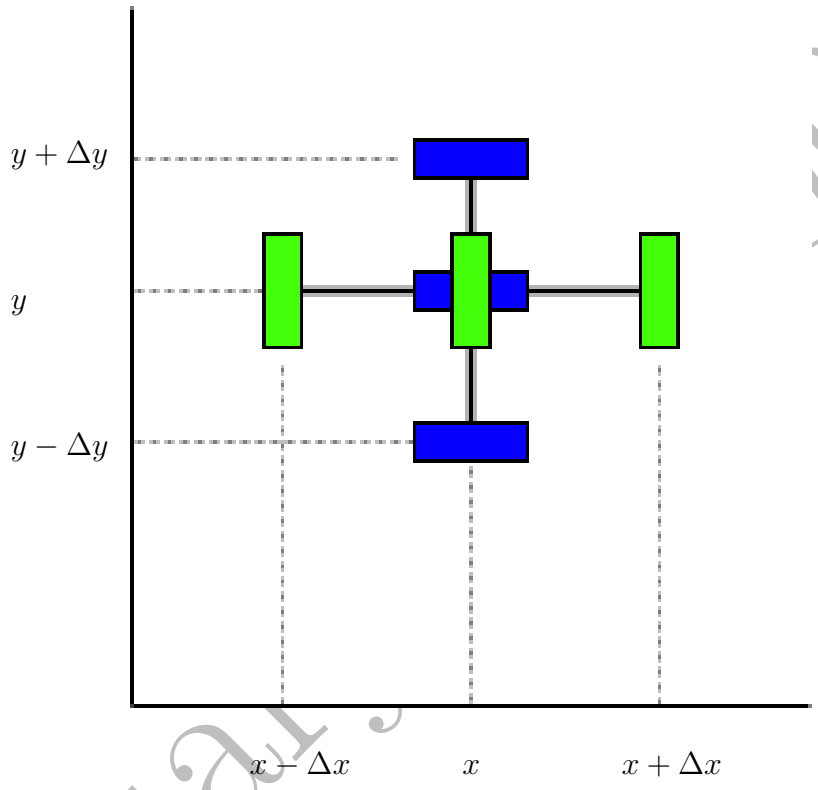


Figure 19.2: Point with error bars

was not considered in the derivation of the equation. It is important that the graph show what really happens so that these unconsidered factors will in fact be noticed and adjusted for. This brings up a second rule: *If the origin is a data point, it is no more "sacred" than any other data point.* In other words, don't force the graph through $(0, 0)$ any more than you would through any other point. Doing a least squares fit will protect you from this temptation.

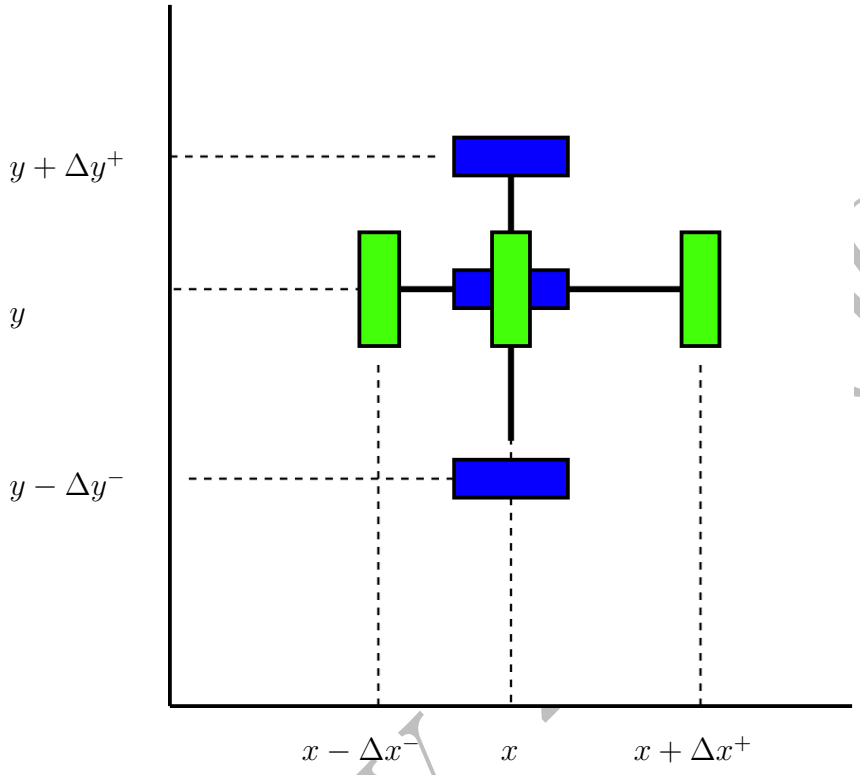


Figure 19.3: Graph with unequal error bars in positive and negative directions

19.3 Graphical Analysis

Usually the point of graphing data is to determine parameters of the mathematical relationship between the two quantities. For instance, when plotting a straight line graph, the slope and y–intercept are the parameters which describe that relationship.

Note that the slope and y–intercept and their uncertainties should have units. The units of the y–intercept should be the same as the y variable, and the slope should have units of

$$[slope] = \frac{[y]}{[x]}$$

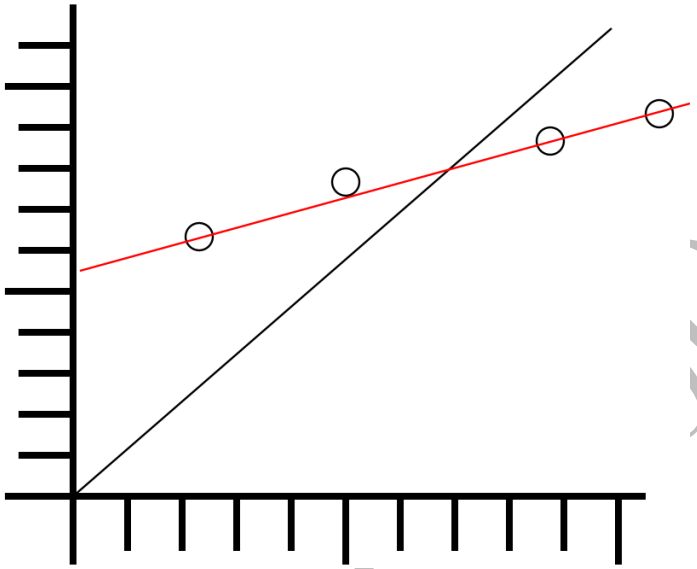


Figure 19.4: A line through the origin is not always the best fit

19.4 Linearizing Equations

In many cases, the mathematical model you are testing will suggest how the data should be plotted. A great deal of simplification is achieved if you can *linearize* your graph, i.e., choose the information to be plotted in such a way as to produce a straight line. (This is discussed in Chapter 20, *"Linearizing Equations"* .) For example, suppose a model suggests that the relationship between two parameters is

$$z = Ke^{-\lambda t}$$

where K and λ are constants. If a graph of the natural logarithm of z is plotted as a function of t, a straight line given by

$$\ln z = \ln K - \lambda t$$

will be obtained. The parameters K and λ will be much easier to determine graphically in such a case.

In particular, if we substitute $y = \ln z$ and $x = t$ in the above equation, and if the slope and y–intercept are measured to be, respectively, m and b, then it should be clear that

$$m = -\lambda$$

and

$$b = \ln K$$

19.5 Curve Fitting

Always draw smooth curves through your data points, unless you have reason to believe that a discontinuity in slope at some point is genuine.

Your graphs should not look like a dot–to–dot drawing.

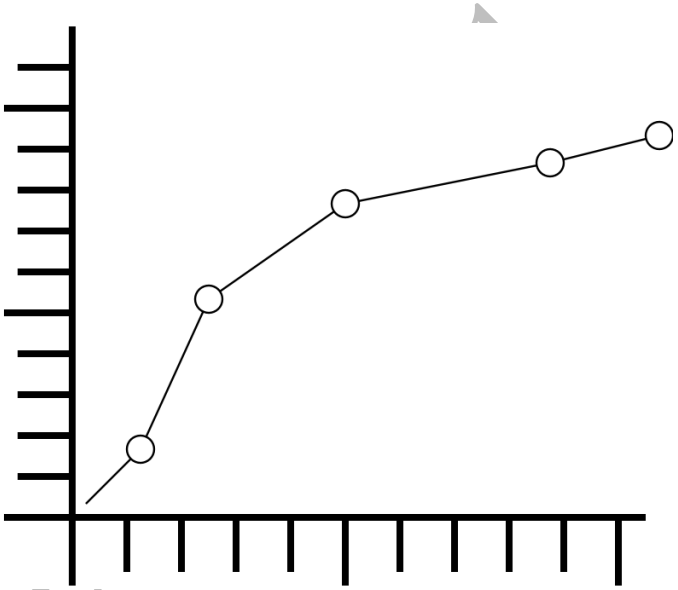


Figure 19.5: Wrong: Graphs should not look like dot-to-dot drawings

If you are plotting the points using the computer, draw the curve by hand if necessary to avoid this problem. However, do not fit data to a curve with no physical significance simply so that all of the points fit.

Do not use an arbitrary function just because it goes through all the data points!

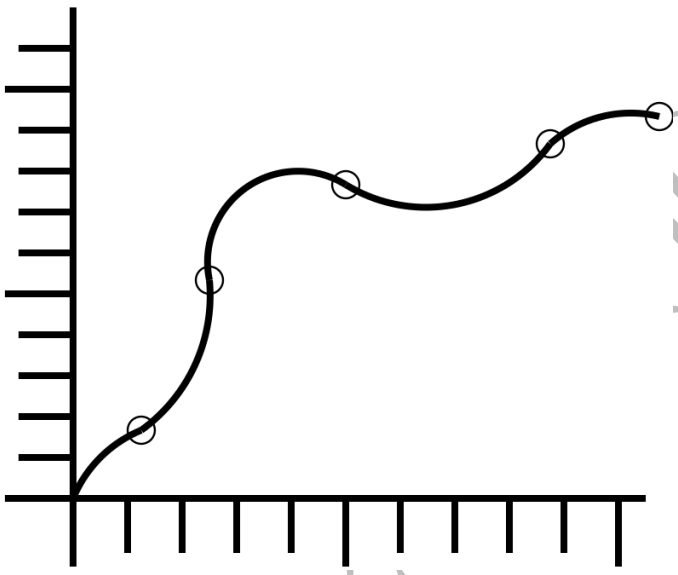


Figure 19.6: Wrong: Graphs should not have meaningless curves just to fit the data

Note that unless a set of data exactly *fits a curve, choosing* a *curve of "best fit" is somewhat arbitrary.* (For example, consider 4 data points at (-1,1), (1,1), (1,-1) and (-1,-1). What line fits these points best?)

Usually, going "by eye" is as good as anything; the advantage to a method such as the *least squares fit* is that it is easily automated, and is generally reliable.

If plotting by eye, one should observe that the line of best fit will usually have an equal number of points above and below it. As well, as a rule, there should not be several points at either end of the graph on the *same* side of the curve. (If this is the case, the curve can be adjusted to avoid this.)

Determining the y-intercept is easy if it is shown on the graph. However if it isn't, you can determine it from the points you used for the slope. If

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

and

$$y = mx + b$$

for any points on the line, including (x_1, y_1) and (x_2, y_2) then

$$y_2 = mx_2 + b$$

so

$$b = y_2 - mx_2$$

and finally

$$b = y_2 - \left(\frac{y_2 - y_1}{x_2 - x_1}\right) x_2$$

19.6 Least Squares Fitting

Least **S**quares **F**itting is a procedure for numerically determining the equation of a curve which "best approximates" the data being plotted. If we wish to fit a straight line to data in the form

$$y = mx + b$$

then the least squares fit gives values for b, the y-intercept, and m, the slope, as follows:[1]

$$b = \frac{\left(\sum y_i\right)\left(\sum x_i^2\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \tag{19.3}$$

and

$$m = \frac{N\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \tag{19.4}$$

(Note: You do not need to calculate uncertainties for m and b during least squares fit calculations like this. Uncertainties in m and b will be dealt with later.)

[1] You may notice that a particular quantity comes up a lot. It is

$$N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2 \tag{19.1}$$

It only takes a couple of lines of algebra to show that this equals

$$N\left(N-1\right){\sigma_x}^2 \tag{19.2}$$

where ${\sigma_x}^2$ is the sample standard deviation of the x values.

If you are interested, an appendix contains a derivation of the least squares fit. In any case you may use the result above.

One important concept which will come up later is that of **degrees of freedom**, which is simply the number which is the difference between the number of data points, (N above), and the number of parameters being determined by the fit, (2 for a straight line). Thus, for a linear fit, the number of degrees of freedom, ν, is given by:

$$\nu = N - 2 \tag{19.5}$$

19.6.1 Correlation coefficient

Equation 19.6 gives the square of the **Pearson product-moment correlation coefficient**, which we will refer to simply as the *correlation coefficient.*[2]

$$R^2 = \frac{\left(N\sum x_i y_i - (\sum x_i)(\sum y_i)\right)^2}{\left(N\sum {x_i}^2 - (\sum x_i)^2\right)\left(N\sum {y_i}^2 - \left(\sum y_i\right)^2\right)} \tag{19.6}$$

The correlation coefficient, R, is a number which has a value between -1 and +1, where a value of -1 indicates a perfect negative correlation, +1 indicates a perfect positive correlation, and a value of zero indicates no correlation. Thus R^2 is a value between zero and 1 indicating just the *strength* of a correlation. The closer R^2 is to one, the stronger the correlation between two variables. To put it another way, the closer it is to one the better one variable can be used as a *predictor* of the other.

19.7 Uncertainties in Graphical Quantities

After the slope and intercept have been calculated, their associated errors are calculated in one of two ways depending on the data. (This is analogous to the idea that the uncertainty in the average is the bigger of the standard deviation of the mean and the uncertainty in the individual values.) The two possible cases are outlined below.

[2] As long as we're dealing with a linear fit, this is the quantity that would commonly be used.

Unless the points fit exactly on a straight line, any graph with big enough error bars will fit the first case, and any graph with small enough error bars will fit the second. It is the relative size of the error bars which determines which case it is.

19.7.1 Small Scatter of Data

If the **scatter** in the data points is small, a straight line which passes through *every* error bar on the graph can be found, as shown in Figure 19.7. This indicates that the uncertainties in your results are *primarily* due to the uncertainties of the measuring instruments used.
The slope and intercept can be found graphically, by eye or using the least squares fit method.

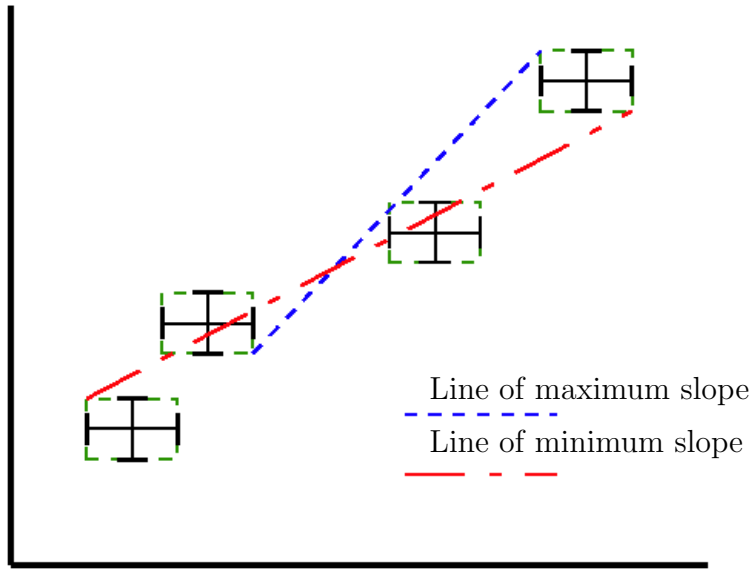


Figure 19.7: Small Scatter of Data Points

To obtain error estimates in these quantities, one draws two lines: a line with the *maximum* slope passing through *all* the error bars; and the line with the *minimum* slope passing through *all* the error bars. These extremes

will determine the required uncertainties in the slope and intercept. *(In this graph and the one following, boxes have been drawn around each point and its error bars to indicate the "uncertainty region" around each point. These would not usually be on a graph, but they are shown here for illustration.)*

The points for the maximum and minimum slope will not always be the endpoints on the graph. Also, the data points providing the endpoints for the two lines will not usually be the same for both. If the maximum and minimum slope are not symmetric about the average, you can calculate

$$\Delta m \approx \frac{m_{max} - m_{min}}{2}$$

and

$$\Delta b \approx \frac{b_{max} - b_{min}}{2}$$

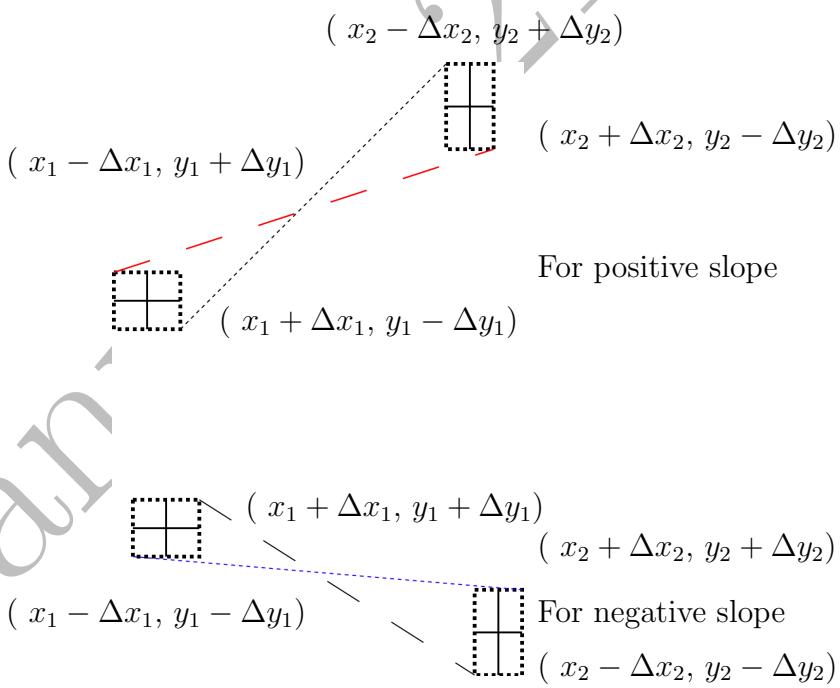


Figure 19.8: Maximum and Minimum Slope Coordinates from a Point

If we label two points x_1 and x_2, where $x_1 < x_2$, then we can see from

Figure 19.8 that the steepest line which touches the error bars for both x_1 and x_2 is the line between $(x_1+\Delta x_1, y_1-\Delta y_1)$ and $(x_2-\Delta x_2, y_2+\Delta y_2)$. The slope of this line will then be

$$m_{max} = \frac{(y_2+\Delta y_2)-(y_1-\Delta y_1)}{(x_2-\Delta x_2)-(x_1+\Delta x_1)} \tag{19.7}$$

and then the y-intercept is given by

$$b_{min} = (y_1-\Delta y_1) - m_{max}(x_1+\Delta x_1) = (y_2+\Delta y_2) - m_{max}(x_2-\Delta x_2) \tag{19.8}$$

Similarly the line with the least slope which touches the error bars for both x_1 and x_2 is the line between $(x_1-\Delta x_1, y_1+\Delta y_1)$ and $(x_2+\Delta x_2, y_2-\Delta y_2)$. The slope of this line will then be

$$m_{min} = \frac{(y_2-\Delta y_2)-(y_1+\Delta y_1)}{(x_2+\Delta x_2)-(x_1-\Delta x_1)} \tag{19.9}$$

and then the y-intercept is given by

$$b_{max} = (y_1+\Delta y_1) - m_{min}(x_1-\Delta x_1) = (y_2-\Delta y_2) - m_{min}(x_2+\Delta x_2) \tag{19.10}$$

The case for a negative slope is shown in Figure 19.8; the analysis is left to the student.

The points for the maximum and minimum slope will not always be the endpoints on the graph.

19.7.2 Large Scatter of Data

Often, you will not be able to find a line which crosses *every* error bar, as with the data in Figure 19.9, and you will have to resort to the numerical method below. In this case, the uncertainties in your graphical results are *primarily* due to the random variations in the data.

Once these values for the slope and intercept are determined, the sum of squares error, S is computed. For the linear case, S can be shown to have a value of

$$S = \sum y_i^2 - m\sum x_i y_i - b\sum y_i \tag{19.11}$$

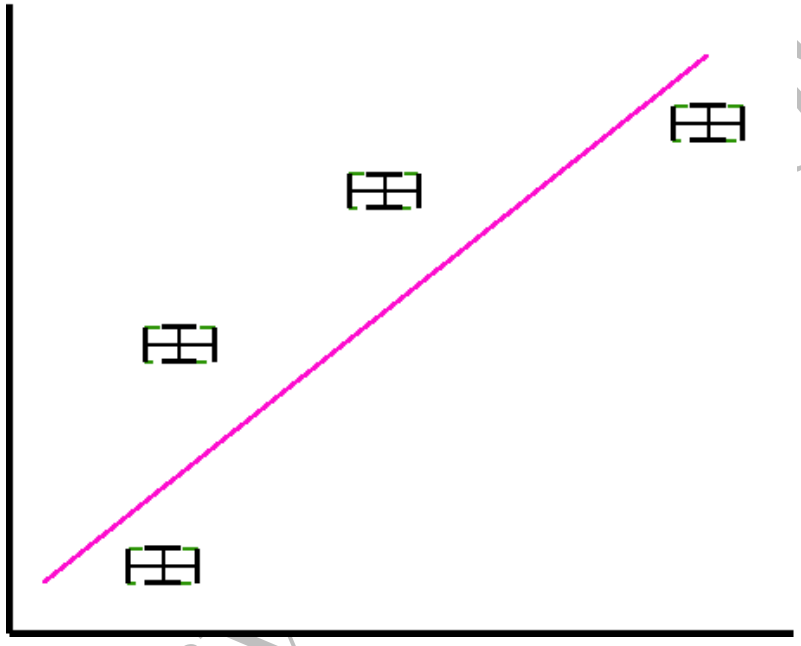


Figure 19.9: Large Scatter of Data

In order to estimate the uncertainty in each parameter, the standard deviation σ is computed from

$$\sigma = \sqrt{\frac{S}{N-2}} \tag{19.12}$$

where $N-2$ is the number of degrees of freedom mentioned earlier. (Often the symbol ν is used for degrees of freedom.) The **standard error** (i.e. uncertainty) in the intercept is

$$\sigma_b = \sigma \sqrt{\frac{\sum x_i^2}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} \tag{19.13}$$

and the standard error (uncertainty) in the slope is

$$\sigma_m = \sigma \sqrt{\frac{N}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} \tag{19.14}$$

i	x_i	x_i^2	$x_i y_i$	y_i	y_i^2
1	0.1	0.01	0.3	3	9
2	0.2	0.04	0.8	4	16
3	0.3	0.09	1.2	4	16
4	0.4	0.16	2.0	5	25
N	$\sum x_i$	$\sum x_i^2$	$\sum x_i y_i$	$\sum y_i$	$\sum y_i^2$
4	1.0	0.3	4.3	16	66

Table 19.2: Sample Least Squares Fit Data

As long as our data fit this second case, then we can use σ_b and σ_m as the uncertainties in the y–intercept and the slope respectively, and use the symbols Δb and Δm instead. Keep in mind, however, that if our data fit the first case, then these terms are not interchangeable. Note that the uncertainties in the slope and y–intercept should have the same units as the slope and y–intercept.

(Note: You do not need to calculate uncertainties for Δm and Δb since these are *uncertainties themselves!)*

19.7.3 Sample Least Squares Calculations

Following is a calculation of the least squares fit and the standard error of the slope and intercept for some test data.

$$N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2 = (4)(0.3) - (1)^2 = 0.2$$

$$b = \frac{\left(\sum y_i\right)\left(\sum x_i^2\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} = \frac{(16)(0.3) - (1)(4.3)}{0.2} = 2.5$$

$$m = \frac{N\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} = \frac{(4)(4.3) - (1)(16)}{0.2} = 6.0$$

$$S = \sum y_i^2 - m\sum x_i y_i - b\sum y_i = (66) - (6)(4.3) - (2.5)(16) = 0.2$$

$$\sigma = \sqrt{\frac{S}{N-2}} = \sqrt{\frac{0.2}{4-2}} = 0.316228$$

$$\sigma_b = \sigma \sqrt{\frac{\sum x_i^2}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} = (0.316228)\sqrt{\frac{0.3}{0.2}} = (0.3878298)$$

$$\sigma_m = \sigma \sqrt{\frac{N}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} = (0.316228)\sqrt{\frac{4}{0.2}} = (1.414214)$$

Thus, if our data are such that σ_b and σ_m are the uncertainties in the y–intercept and the slope, and thus Δb and Δm, then

$$b = 2.5 \pm 0.4$$

and

$$m = 6 \pm 1$$

19.8 References

- *The Analysis of Physical Measurements*, Emerson M. Pugh and George H. Winslow, Addison-Wesley Series in Physics, 1966, QC39.P8
- *Errors of Observation and Their Treatment*, J. Topping, Chapman and Hall Science Paperbacks, 1972(4th Ed.)
- *Statistics*, Murray R. Speigel, Schaum's Outline Series in Mathematics, McGraw-Hill, 1961

Chapter 20

Linearizing Equations

20.1 Theory

Often, the point of a scientific experiment is to try and find empirical values for one or more physical quantities, given measurements of some other quantities and some mathematical relationship between them. For instance, given a marble has a mass of 5 g, and a radius of 0.7 cm, the density of the marble can be calculated given that $v = 4/3\pi r^3$ and $\rho = m/v$. (For the sake of simplicity, uncertainties will be ignored for now, although the calculation of those should be familiar by now.)

Many times, however, rather than having *one* measurement of a quantity, or set of quantities, we may have several measurements which should all follow the same relationships, (such as if we had several marbles made of the same material in the example above), and we wish to combine the results. The usual way of combining results is to create a graph, and extract information (such as the density) from the slope and y–intercept of the graph.

One may be tempted to ask why a graph should be better than merely averaging all of the data points. The answer is that an average is completely unbiased. The variation of any one point from the norm is no more or less important than the variation of any other point. A graph, however, will show any point which differs significantly from the general trend. Analysis of the graphical data (such as with a least squares fit) will allow such "outliers" to be given either more or less weight than the rest of the data as the researcher deems appropriate. Depending on the situation, the researcher may wish to

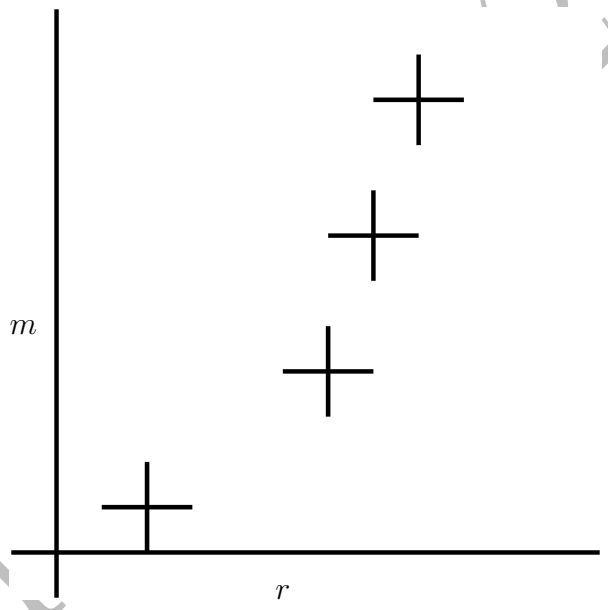


Figure 20.1: Non-linear equation

verify any odd point(s), or perhaps the trend will indicate that a linear model is insufficient. In any case, it is this added interpretive value that a graph has which makes it preferable.

A plot is better than an average since it may indicate systematic errors in the data.

The value in fitting the data to an equation is that once the fit has been done, rather than continuing to work with a large amount of data, we can simply work with the parameters of our fit and their uncertainties. In the case of a straight line, all of our data can be replaced by four quantities; $m, \Delta m, b$ and Δb.

A fit equation replaces a bunch of data with a few parameters.

The reason a linear graph is so useful is that it's easier to identify whether a line is straight than it is to identify whether it looks more like $y = x^2$ or $y = x^3$, for instance.

A straight line is easy to spot with the unaided eye.

If the data fits an equation of the form $y = mx + b$, then it is easy to plot a straight line graph and interpret the slope and y–intercept, but it is rarely that simple. In most cases, the equation must be modified or **linearized** so that the variables *plotted* are different than the variables *measured* but produce a straight line.

Linearizing equations is this process of modifying an equation to produce new variables which can be plotted to produce a straight line graph. In many of your labs, this has been done already.

Look again at $y = mx + b$. Note that y and x are *variables*, (as each can take on a range of values), while m and b are *constants*, (as there is only one value for each for all of the data points). We can linearize an equation if we can get it in the form

$$variable_1 = constant_1 \times variable_2 + constant_2$$

There are a few things to note:

1. Several constants combined together produces another single constant.

2. Powers or functions of constants are also constants.

3. Constants may have "special" values of 0 or 1 so they appear "invisible". For example

$$y = mx$$

is still the equation of a straight line, where $b = 0$. As well,

$$y = b$$

is the equation of a line where $m = 0$.

4. Variables may be combined together to form new variables.

5. Powers or functions of variables are also variables.

Note that linearizing an equation will produce expressions for the slope and y–intercept which depend only on the constants in the original equation, not *on the original x and y variables. This means that the constants can be related to the slope and y–intercept rather than the original variables.*

20.1.1 Techniques for Linearization

If a relationship involves *only* multiplication and division, (including powers), then logarithms can be used to linearize. Sometimes taking roots or powers of both sides of an equation will help.

20.1.2 Procedure for Linearization

The steps are as follows:

1. Rearrange the equation to get one variable (or a function of it) on the left side of the equation; this becomes your y variable.

2. Regroup the right side of the equation to create a term containing the other variable (or some function of it).

3. Use the left-side variable (or the function of it) as your x variable, and then your slope should be whatever multiplies it; your y intercept is whatever additive term is left over.

Note: It is important to realize that you don't need to understand an equation to linearize it; all you have to know is which parameters are variables (ie. things you have data for), and which parameters are constants (ie. things you want to calculate). Of course different experiments involving the same relationship may make different parameters variable, and so how an equation is linearized will depend on the data used. To again consider the above example: The original equations were

$$v = (4/3)\pi r^3 \tag{20.1}$$

and

$$\rho = m/v \tag{20.2}$$

where the quantities m and r are measured. (ie. We have several marbles of the same material, so we can get several measurements of m and r, but we expect ρ to be the same for all of them.) Thus for this situation, m and r are variables, and ρ is a constant. We can combine the two equations to get

$$\rho = \frac{m}{(4/3)\pi r^3} \tag{20.3}$$

or

$$\rho = \frac{3m}{4\pi r^3} \tag{20.4}$$

This equation has a constant on one side, and a mixture of variables and constants on the other. First we should rearrange it to get a variable on the left hand side. Suppose we rearrange the equation, giving

$$m = (4/3)\pi\rho r^3 \tag{20.5}$$

This leaves a variable on the left. From this point on, there are two main possibilities for how to proceed: [1]

Method I

Now we can create a new variable, Y such that

$$Y = m$$

[1] Usually the process is not as explicit as this. ie. one doesn't usually create an X and a Y, but doing this illustrates the procedure.

By the rule about powers of variables being variables, then we can create a new variable X given by

$$X = r^3$$

Then equation 20.5 above becomes

$$Y = (4/3)\pi\rho X \tag{20.6}$$

since π is a constant, and ρ *should* be, and using the rule that combinations of constants produce constants, then we can define M, a constant, (not the same as m), as

$$M = (4/3)\pi\rho$$

so equation 20.6 becomes

$$Y = MX + 0$$

which is the equation of a straight line. (In the case, B, the y–intercept is zero.)[2] So if we plot our "modified" variables, we should get a straight line, passing through the origin with a slope M. How can we get ρ from the graph? Well, from above

$$M = (4/3)\pi\rho$$

so

$$\rho = \frac{3M}{4\pi}$$

where M is the slope of the graph.

Method II

We can take logarithms of both sides, so that Y such that equation 20.5 above becomes

$$\ln m = \ln\left((4/3)\pi\rho\right) + \ln r^3 \tag{20.7}$$

grouping the terms so one only contains constants (and so the combination should be constant) and one only contains the variable r. We can bring down the exponent so equation 20.7 becomes

$$\ln m = \ln\left((4/3)\pi\rho\right) + 3\ln r$$

[2]Occasionally we can get a situation where the slope is similarly "invisible", if it is 1 or 0.

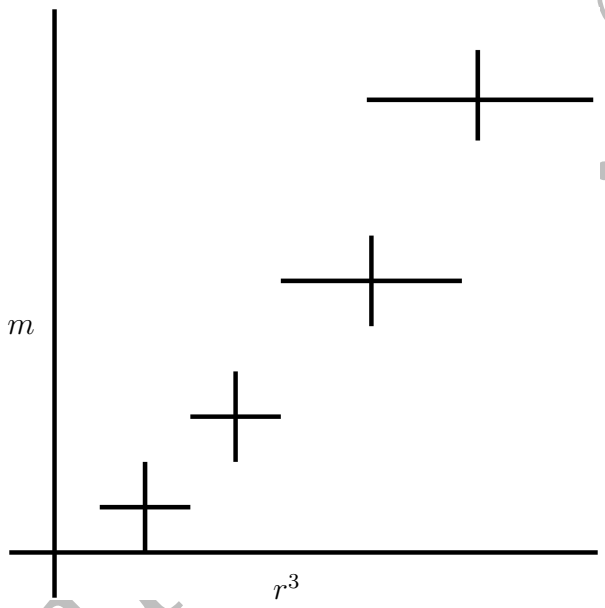


Figure 20.2: One linearization

Now we can create new variables, Y such that

$$Y = \ln m$$

and

$$X = \ln r$$

which is the equation of a straight line. So if we plot our "modified" variables, we should get a straight line. How can we get ρ from the graph? Well, from above

$$B = \ln\left((4/3)\pi\rho\right)$$

so

$$\rho = \frac{3}{4\pi}e^{B}$$

where B is the y-intercept of the graph. (In this case, the value you get from the graph for the slope should suggest whether the fit is a good one.)

Remember that after linearization, our results depend on our graphical quantities of the slope and the y-intercept, rather than on the original measured quantities.

20.1.3 Choosing a Particular Linearization

Often there may be more than one linear form for the equation so there may be more than one "right answer". In this case, there are a few things which may help you choose.

Simple variables

A preferable linearization is one which most simplifies understanding the graph or interpreting the results. For instance, in the above example, it would have been possible to use $(4/3)\pi r^3$ instead of r^3 as our x variable, but that would make confusing axis scales and/or units (although it would have made the slope be ρ with no calculation).

Spread of data

The spread of data will be different for each linearization. A graph with points which are more equally spaced is generally preferable to one where the points are concentrated in one area.

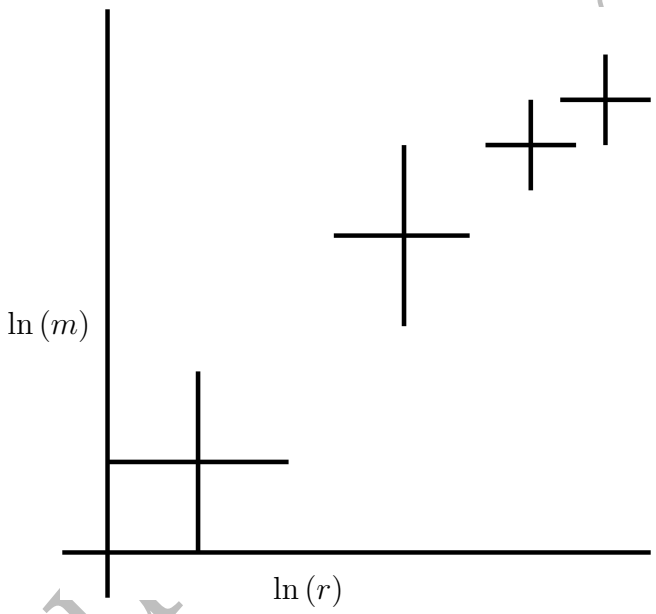


Figure 20.3: Another linearization

Size of error bars

Like the spread of data, the size of the error bars will be different for each linearization. A graph with more equally sized error bars is generally preferable to one where the error bars vary greatly in size for different points.

Usually it is preferable to separate variables and constants as much as possible in your linearization so that graph variables are easily related to experimental ones.

20.1.4 Uncertainties in Results

After determining how equation parameters relate to graphical quantities as above, uncertainties can be determined as usual. In the above example Method I gives

$$\Delta\rho = \frac{3\Delta M}{4\pi}$$

while for Method II

$$\Delta\rho = \frac{3}{4\pi}e^B\Delta B$$

or

$$\Delta\rho = \rho\Delta B$$

For instance, from the example above, using Method I we would say:

- Plot m vs. r^3. (In other words, the independent (x) variable is r^3 and dependent (y) variable is m.)
- The uncertainty in the dependent variable is Δm.
- The uncertainty in the independent variable is $3r^2\Delta r$.
- The slope of the graph will be $M = (4/3)\pi\rho$.
- The y-intercept should be zero[3].
- The density will be determined from the slope by the equation $\rho = \frac{3M}{4\pi}$.
- The uncertainty in the density will be determined from the slope by the equation $\Delta\rho = \frac{3\Delta M}{4\pi}$.

[3]If the y-intercept turns out to be something other than zero, then there is some systematic error in our experiment.

20.2 Recap

By the end of this exercise, you should understand the following terms:

- linear graph
- linearized equation

Chapter 21

Lab Reports: General Outline

A lab report is *personal*, in the sense that it explains what *you* did in the lab and summarizes *your* results, as opposed to an assignment which generally answers a question of some sort. On an assignment, there is (usually) a "right answer", and finding it is the main part of the exercise. In a lab report, rather than determining an "answer", you may be asked to *test* something. (Note that no experiment can ever *prove* anything; it can only provide evidence for or against; just like in mathematics finding a single case in which a theorem holds true does not prove it, although a single case in which it does not hold refutes the theorem. A **law** in physics is simply a theorem which has been tested countless times without evidence of a case in which it does not hold.) The point of the lab report, when testing a theorem or law, is to explain whether or not you were successful, and to give reasons why or why not. In the case where you are to produce an "answer", (such as a value for g), your answer is likely to be different from that of anyone else; your job is to describe how you arrived at yours and why it is reasonable under the circumstances.

21.1 Format of a Lab Report

The format of the report should be as follows:

21.1.1 Title

The title should be more specific than what is given in the manual; it should reflect some specifics of the experiment.

21.1.2 Purpose

The specific purpose of the experiment should be briefly stated. (Note that this is not the same as the goals of the whole set of labs; while the labs as a group are to teach data analysis techniques, etc., the specific purpose of one experiment may indeed be to determine a value for g, for instance.) *Usually, the purpose of each experiment will be given in the lab manual. However, it will be very general. As in the title, you should try and be a bit more specific.*

There should always be both **qualitative** and **quantitative** goals for a lab.

Qualitative

This would include things like *"see if the effects of friction can be observed"*. In order to achieve this, however, specific quantitative analyses will need to be performed.

Quantitative

In a scientific experiment, there will always be numerical results produced which are compared with each other or to other values. It is based on the results of these comparrisons that the qualitative interpretations will be made.

21.1.3 Introduction

In general, in this course, you will not have to write an Introduction section.

An introduction contains two things: *theory* for the experiment and *rationale* for the experiment.

Theory

Background and theoretical details should go here. Normally, detailed derivations of mathematical relationships should not be included, but references must be listed. All statements, equations, and 'accepted' values must be justified by either specifying the reference(s) or by derivation if the equation(s) cannot be found in a reference.

Rationale

This describes why the experiment is being done, which may include references to previous research, or a discussion of why the results are important in a broader context.

21.1.4 Procedure

The procedure used *should not be described* unless you deviate from that outlined in the manual, or unless some procedural problem occurred, which must be mentioned. A reference to the appropriate chapter(s) of the lab manual is sufficient most of the time.

Ideally, someone reading your report and having access to the lab manual should be able to *reproduce* your results, within reasonable limits. (Later on we will discuss what "reasonable limits" are.) If you have made a mistake in doing the experiment, then your report should make it possible for someone else to do the experiment *without* making the same mistake. For this reason, lab reports are required to contain **raw data**, (which will be discussed later), and **explanatory notes**.

Explanatory notes are recorded to

- explain any changes to the procedure from that recorded in the lab manual,
- draw attention to measurements of parameters, values of constants, etc. used in calculations, and
- clarify any points about what was done which may otherwise be ambiguous.

Although the procedure need not be included, your report should be clear enough that the reader does not need the manual to understand your write–up.

(If you actually need to describe completely how the experiment was done, then it would be better to call it a "Methods" section, to be consistent with scientific papers.)

21.1.5 Experimental Results

There are two main components to this section; *raw data* and *calculations*.

Raw Data

In this part, the reporting should be done part by part with the numbering and titling of the parts arranged in the same order as they appear in the manual.

The *raw data* are provided so that someone can work from the actual numbers you wrote down originally before doing calculations. Often mistakes in calculation can be recognized and corrected after the fact by looking at the raw data.

In this section:

- Measurements and the names and precision measures of all instruments used should be recorded; in tabular form where applicable.
- If the realistic uncertainty in any quantity is bigger than the precision measure of the instrument involved, then the cause of the uncertainty and a bound on its value should be given.
- Comments, implicitly or explicitly asked for regarding data, or experimental factors should be noted here. This will include the answering of any given *in-lab* questions.

Calculations

There should be a clear path for a reader from raw data to the final results presented in a lab report. In this section of the report:

- Data which is modified from the original should be recorded here; in tabular form where applicable.
- Uncertainties should be calculated for all results, unless otherwise specified. The measurement uncertainties used in the calculations should be those listed as realistic in the raw data section.
- Calculations of quantities and comparisons with known relationships should be given. If, however, the calculations are repetitive, only one sample calculation, shown in detail, need be given. Error analysis should appear here as well.

- Any required graphs would appear in this part. (More instruction about how graphs should be presented will be given later.)
- For any graph, a table should be given which has columns for the data (including uncertainties) which are actually plotted on the graph.
- Comments, implicitly or explicitly asked for regarding calculations, observations or graphs, should be made here.

Sample calculations may be required in a particular order or not. If the order is not specified, it makes sense to do them in the order in which the calculations would be done in the experiment. If the same data can be carried through the whole set of calculations, that would be a good choice to illustrate what is happening.

Printing out a spreadsheet with formulas shown does not count as showing your calculations; the reader does not have to be familiar with spreadsheet syntax to make sense of results.

Post-lab questions should not be answered ina numbered list; rather the answers should be integrated in to the *Discussion* and *Conclusion* sections based on where they would be most appropriate.

21.1.6 Discussion

This section is where you explain the significance of what you have determined and outline the *reasonable limits* which you place on your results. (This is what separates a scientific report from an advertisement.) It should outline the major sources of random and systematic error in an experiment. *Your emphasis should be on those which are most significant, and on which you can easily place a numerical value. Wherever possible, you should try to suggest evidence as to why these may have affected your results, and include recommendations for how their effects may be minimized.* This can be accompanied by suggested improvements to the experiment.

Two extremes in tone of the discussion should be avoided: the first is the "sales pitch" or advertisement mentioned above, and the other is the "apology" or disclaimer (*" I wouldn't trust these results if I were you; they're probably hogwash."*) *Avoid whining* about the equipment, the time, etc. Your

job is to explain briefly what factors most influenced your results, not to absolve yourself of responsibility for what you got, but to suggest changes or improvements for someone attempting the same experiment in the future. Emphasis should be placed on improving the experiment by changed *technique*, (which may be somewhat under your control), rather than by changed *equipment*, (which may not).

*M*any of the in-lab questions are directed to things which ought to be discussed here. Like the post-lab questions, don't answer them in a list, but integrate them into the text.

This section is usually worth a large part of the mark for a lab so be prepared to spend enough time thinking to do a reasonable job of it.

You must discuss at least one source of systematic error in your report, even if you reject it as insignificant, in order to indicate how it would affect the results.

21.1.7 Conclusions

Just as there are always both **qualitative** and **quantitative** goals for a lab, there should always be both **qualitative** and **quantitative** *conclusions* from a lab.

Qualitative

This would include things like *"see if the effects of friction can be observed"*. In order to achieve this, however, specific quantitative analyses will need to be performed.

Quantitative

In a scientific experiment, there will always be numerical results produced which are compared with each other or to other values. It is based on the results of these comparrisons that the qualitative interpretations will be made.

General comments regarding the nature of results and the validity of relationships used would be given in this section. Keep in mind that these comments can be made with certainty based on the results of error calculations.

The results of the different exercises should be commented on individually. *Your conclusions should refer to your original purpose; eg. if you set out to determine a value for g then your conclusions should include your calculated value of g and a comparison of your value with what you would expect.*

While you may not have as much to say in this section, what you say should be clear and concise.

21.1.8 References

If an 'accepted' value is used in your report, then the value should be footnoted and the reference given in standard form. Any references used for the theory should be listed here as well.

21.2 Final Remarks

Reports should be clear, concise, and easy to read. Messy, unorganized papers never fail to insult the reader (normally the marker) and your grade will reflect this. A professional report, in quality and detail, is at least as important as careful experimental technique and analysis.

Lab reports should usually be typed so that everything is neat and organized. Be sure to spell check and watch for mistakes due to using words which are correctly spelled but inappropriate.

21.3 Note on Lab Exercises

Lab *exercises* are different than lab reports, and so the format of the write-up is different. Generally exercises will be shorter, and they will not include either a *Discussion* or a *Conclusion* section.

Computer lab exercises may require little or even no report, but will have points which must be demonstrated in the lab.

Chapter 22

Lab Reports: Finer Points

22.1 Introduction

Previously, you have prepared lab reports primarily with a *marker* in mind. However you should be starting to gear your reports to a more general *reader*. How this will change your report should become clear as you read the following:

22.1.1 Purpose of a report

The goal of presenting a report is to *inform*, not to *impress*. That means that, on the one hand, you don't want to fill space with drivel just to make the reader think you know something, (it's not likely to work), but on the other hand, at times it may be helpful to repeat a piece of useful information two or three times in a report to save the reader having to flip back and forth. Individual sections should be as self–contained as possible, so that a reader is not normally forced to hunt for pertinent facts all through the report.

22.1.2 Structuring a report

In some labs, lab templates may have been used to organize reports in a very standard way to give some uniformity to the reports. Now you will not have that order imposed, and so you will have to structure your own reports so that they are understandable. Part of what this will require is for you to put in enough *"English glue"* to make the report easy to read, even (especially!) for someone who does not have the lab manual at hand.

22.1.3 Moving from lab notes to a report

When handling data, either to analyze it or to present it, it is important to make a distinction between between *utility* and *clarity*. In other words, how you set up a spreadsheet to analyze data or graph it may *not* be the way that you should set it up for someone else to look at. Similarly, showing the output block from a least squares fit reflects whether you did it correctly, but what you should present is the *meaningful* results of the fit, not every bit of output. (If it needs to be included for a marker, put it in an appendix so that it's there, but does not hurt the flow of the report.) Following are some guidelines for presenting data for the *reader*, not for the *writer* or the *marker*.

22.2 Text

- Grammar and spelling count!
- All numerical quantities must include uncertainties!
- In the text of a report, all symbols should be explained, especially if they are non–standard (for instance if you use "w" instead of "ω"). For instance *"w is the angular frequency in rads/sec.")* Units should be given for each quantity as well.
- The report should have brief descriptions of procedures, etc., so that a person not following the manual can still make sense of the data. If you are following a manual, you need not go into great detail, but the significance of parameters stated, etc. should be explained; eg. *"current was measured by calculating the voltage across resistor R_M"*
- Quotes, standard values, etc. should be foot-noted and referenced.
- Derivations may be done by hand (if long), but if you are using a word processor, this is a chance to learn more features if you use it to do at least the short derivations. Make sure the symbols you use in derivations match the symbols you use in the text. (See above example with w and ω.)
- Watch for similar or duplicate symbols; eg. e, the base for natural logarithms, and e the charge on an electron (or in the above example

where you might have both w and ω). If you have to use two symbols like this in one report, change one (and define it!) or change both. (For example, use q for the charge on an electron, or replace e^x with $\exp(x)$.)

- Quantities which would normally be expressed using scientific notation with uncertainties should usually be presented in the standard form; eg. $(2.3 \pm 0.2) \times 10^{-6} m$.

The purpose of scientific notation is to remove **placeholder** *zeroes, either before or after the decimal point, from a number. Thus it often does NOT make sense in numbers in the range from about* $0.1 \rightarrow 99$*, where there are no placeholder zeroes.*

- All results should be given the correct number of significant figures; ie. one or at most two significant figures for uncertainties, and quantities rounded so least significant digit is in the same place as the least significant digit of uncertainty.

22.3 Tables

- Always include tables of *raw data*, even if you need to modify the data to plot a graph. That way if you make a mistake in calculations, it will be possible to correct later.
- Tables must have boxes around them and lines separating columns, etc. ie. unstructured spreadsheets are *not* OK.
- Any data which will be plotted in a graph should be shown in a table with the same units and uncertainties as on the graph.
- No table should be split by a page break; if necessary make it into two separate tables.
- All tables need names and numbers such as "Table 1" (which should be referenced in the report), and *meaningful* labels which match the text (or explanations of how the labels correspond to the quantities in the text). Table 22.1has two obvious problems; the column labels are somewhat cryptic, and much data is redundant.

Vx	dVx	Ix	dIx	Used (y/n)
0.10	0.02	1.5	0.3	y
0.19	0.02	1.3	0.3	y
0.24	0.02	1.2	0.3	y
0.41	0.02	0.8	0.3	n

Table 22.1: One way to do it

In Table 22.2, some of this is changed. This is much less "busy", and more descriptive.

Voltage	Current
0.10	1.5
0.19	1.3
0.24	1.2
0.41	$0.8^{\dagger}$

All voltages $\pm$ 0.02 volts.
All currents $\pm$ 0.3 amps.
† point not used in fit.

Table 22.2: More concise way

Least squares fit output can be somewhat confusing; indicate which points were used (if not all, as in the above example), the fit equation, and the parameters calculated as well as their standard errors. Be sure to include the proper units for both. A table may not even be a good way to give these.

22.4 Graphs

- *All graphs must include error bars!* If error bars in one or both dimensions are too small to be seen on a graph, then a note should be made on the graph to indicate this.

- Titles should be *descriptive*; ie. they should give pertinent information which is not elsewhere on the graph.

- Graphs can be annotated with fit results so that by looking at the graph the reader can see fit results (with uncertainties, of course). Make sure

to mark which points were used for the fit (if not all). Keep in mind the above rules about symbols.

- Fit results should be given in terms of actual graphical quantities, not x and y. For instance, *"Slope is* 4.5 ± 0.2 *N/kg"; y-intercept is* $3.1 \pm 0.1N$*"*, as opposed to *"y=4.5x+3.1"* which lacks uncertainties, units, and *relevance*.

- Put units on each axis, and either use a grid for both dimensions, or else none at all. (A horizontal–only grid looks kind of odd.)

- Graphs should be in the specified orientation. *NOTE: a graph of y versus x means y is on the vertical axis and x is on the horizontal axis.*

22.5 Linearizations

- Always include the original (ie. non-linearized) equation(s) as well as the linearized one(s).

22.6 Least Squares Fits

- Always plot data (with error bars) before fitting to see that points make sense. (Make sure error bars are correct!)

- Clearly identify data used in fit (if not all points are used).

- Give results meaningful names, such as "slope", "standard error in slope", etc.

- Include *units* for slope, y-intercept, etc.

- Show the fit line on the graph with the data.

- Identify whether the graph shows "small" or "large" scatter, and then according to that identification, do whichever of these is appropriate:

  – Perform fit in such a way as to get standard errors in both y-intercept and slope.

  - Find maximum and minimum slopes (if they exist) and show them on the graph with the data.

- Determine the uncertainties in the slope and the y–intercept from the result above.
- If you would have expected either the slope and the y–intercept to be zero, and it isn't, then suggest why that might be so.

22.7 Other

Printing out a spreadsheet with formulas shown does not count as showing your calculations; the reader should not have to be familiar with spreadsheet syntax to make sense of results.

You must discuss at least one source of systematic error in your report, even if you reject it as insignificant, in order to indicate how it would affect the results.

Chapter 23

Introduction to Spreadsheets

23.1 Theory

Some of the information here will be specific to Microsoft Excel, but most should be similar for other spreadsheets as well.

23.1.1 Formulas

The value in using *cell references* instead of numerical values is that if the numbers change, the calculations are performed automatically. In a lab this allows you to set up the spreadsheet and then simply type in new data to see new results. *If you do things properly*, it is even easy to change the number of data points after the fact.

Never use a number where you can use a cell reference.

23.1.2 Functions

Functions perform commonly used tasks on a cell or block of cells and return the result in a cell.[1] Some of the most common ones follow.

Common Functions

For blocks of cells

[1] There are functions which return a block of cells, which don't fit with the mathematical notion of a "function", but we won't get into those here.

- **average()**
- **count()** counts the non-empty cells

Always use the count function instead of typing in the number of data points so you can change the number later if needed without editing formulas.

- **stdeva()** determines the sample standard deviation
- **max()** finds the maximum
- **min()** finds the minimum

For individual values

- **sqrt()** square root
- **abs()** absolute value
- **if()** allows a cell to have a value depending on a condition

Using the Built-in Automation

In you aren't sure what parameters are needed by a function, or in what order, you can simply type in the function until after the left bracket and then you will be prompted for what needs to be filled in.

23.1.3 Copying Formulas and Functions: Absolute and Relative References

Never copy values from one part of the spreadsheet to another where a cell reference could be used instead.

Relative References

The value of using formulas in a spreadsheet is that calculations can be easily repeated. One of the ways this happens is by copying formulas to be applied to different data. When a formula is copied from one location to another, references are usually changed *relative* to the move. In other words, if a

formula is copied from one cell to the next one on the right (ie. where the column address increases by one), then cell references *within* the formula will have their column addresses increased by one as well. If a formula is copied from one cell to the one above (ie. where the row address decreases by one), then cell references within the formula will have their *row* addresses decreased by one as well.[2]

Figure 23.1 gives an example of both absolute and relative references and how they affect copying. If the formula given was in the coloured cell, and then copied to the cell shown, the references in the formula would become the ones indicated.

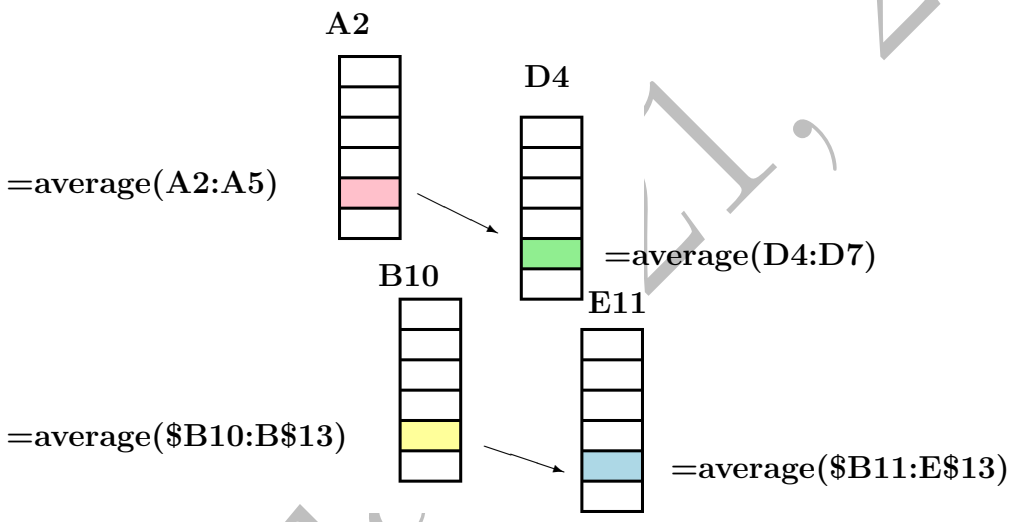


Figure 23.1: Effects of absolute and relative references when copying

Absolute References

Sometimes you don't want cell references in a formula to change when it is copied. In this case what you do is to put a **$** before the part of a cell reference which you do not want to change. in other words, if you have a formula with a reference to **B13** and you do not want the **B** to change when you copy, then change the reference to **$B13**. If you don't want the **13** to change when you copy, then change the reference to **B$13**. If you don't want

[2]Note that when you *move* a cell, none of the references inside it are changed.

either **B** or **13** to change, then change the reference to **B13**. (In other words, you make a number a constant by referring to it like **B13**).

23.1.4 Pasting Options

Sometimes when you copy a formula, you don't want to copy everything about it. For instance, sometimes you just want to get the numerical result. In a spreadsheet, you can't choose what to copy, but you can choose what to *paste*. When you go to paste something, you can **Paste Special** and choose to paste whatever you want. Thus you can paste strings or numbers and not formulas, and you can include formatting (boxes, colours, fonts, etc.) or not.

Always copy formulas from one part of the spreadsheet to another instead of retyping to avoid making mistakes.

23.1.5 Formatting

You can do many things to a cell or block of cells to change its appearance, such as putting boxes around it, changing fonts or how numbers are displayed, changing the background colour, etc.

Format after you're done copying formulas, etc., to avoid having to always Paste Special *to preserve style information.*

23.1.6 Print Preview

When printing a spreadsheet, there are lots of options to make the result more readable. Doing a *print preview* allows this. You can

- turn the grid off.
- include page headers and footers (or not)
- scale to fit the page

Doing a page preview also helps you figure out which page number(s) you need to print, since it may not be obvious.

Preview the printing to avoid wasting reams of paper printing out stuff you don't want.

23.2 Recap

By the end of this exercise, you should know how to use spreadsheet functions to calculate the :

- mean
- standard deviation
- standard deviation of the mean

Chapter 24

Graphing and Least Squares Fitting in Excel

24.1 Theory

24.1.1 Graphing

Graph Type

The graph type mainly used by scientists is an **x-y** graph[1]. *Do not* choose a line graph!

Colour

The default grey background in many spreadsheets just looks bad in graphs; it obscures the data and serves no purpose. *Turn the background colour off!*

Gridlines

Grid lines should be either removed or in both dimensions. Gridlines in one direction only look odd on an "xy" graph. *Turn the gridlines off!*

[1] As long as there is some *mathematical relationship* between the variables, then an x-y graph illustrates the relationship. However, if the independent variable does not have a numerical value, then this doesn't apply. For instance, if you were graphing reaction time for men and women, then a bar graph would be the logical choice, since there's no *numerical* relationship between "men" and "women".

Text

The main text of a graph consists of $x-$ and $y-$ titles, a *main title* and perhaps a *sub–title*. All of these may be set in Excel.

Series

Excel allows you to plot several different "series" of (x, y) data. Each series can be *customized*, with choices for many things, including the following:

- Patterns

  Each series can be plotted with lines, symbols, or both.

Do not connect the points like a dot-to-dot drawing!

Do not use an arbitrary function just because it goes through all the data points!

- Markers

  There are many possible symbols which can be used for each series.

- Lines

  There are several line types available for each series.One important fact about how lines are used to connect points in a series; *all points in a series are joined by lines*, unless the line for that series is turned off.

In science, it is almost always wrong to have a dot-to-dot drawing. It is also wrong to have a curve which has no mathematical significance. For this reason, data points should not be connected by either line segments or a curve like a polynomial which is made to pass through each data point. The only line or curve which should be shown is the result of a fit which is based on some theoretical mathematical relationship.

Matching up x and y Values

When you create an xy graph is created in Excel, you *don't* input data values as (x, y) pairs. Instead you select series for each of x and y. The way the

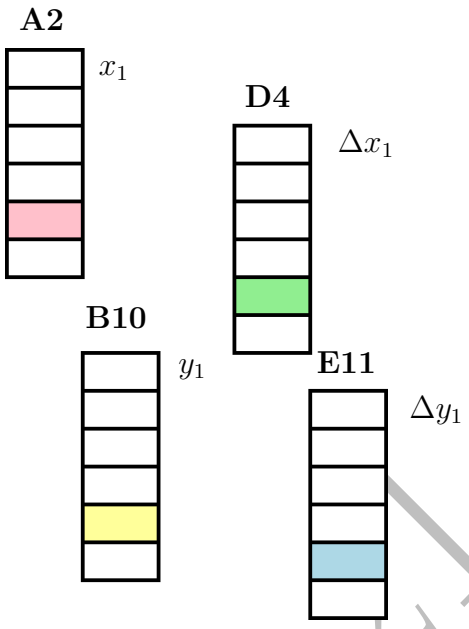


Figure 24.1: Spreadsheet layout with series for x, y, and error bars

individual x and y values are associated is by where they occur in their respective series.

In Figure 24.1 you can see that the 5^{th} point in each series is highlighted. Even though the series all start in different rows and columns, since the number of cells in each is the same, corresponding values can be considered to be related. (If a cell is blank, then the corresponding point or error bar will not be plotted.)

Error Bars

In Excel, when you choose *custom* error bars, you can choose series for both x and y, and even potentially different series for the + and − directions.

You don't necessarily have to put markers on the ends of the error bars; the value in doing so is to make it clear that you're not just using "+" symbols for plotting the data points. Also, if you are including a grid on your graph, error bars without markers on the end may be hard to distinguish. However if the error bars will be clearly identifiable without markers, you don't need to use them.

24.1.2 Least Squares Fitting

The point of plotting a graph in an experiment is usually to extract information from the graph; often the data is plotted in such a manner that the model being tested suggests that the data should fit a straight line. If it does, then getting the slope and y intercept of the line of best fit along with their associated uncertainties is necessary. One of the two usual ways to determine the uncertainty in a graphical quantity is to calculate the **standard error**. (The other involves finding lines of maximum and minimum slope.) The following sections discuss using Excel to do least squares fitting and to calculate standard errors.

Determining the equation of the line by formulas

In Chapter 19, *"Graphs and Graphical Analysis"*, the lab manual explains how to calculate a least squares fit to a set of data. This can be done in Excel by creating additional cells corresponding to each data point which contain, respectively, x^2, y^2 and xy. At the end of the data, these quantities can be totaled to give the sums necessary to do the least squares fit.[2]

[2]You may notice that a particular quantity comes up a lot. It is

$$N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2 \tag{24.1}$$

It only takes a couple of lines of algebra to show that this equals

$$N\left(N-1\right){\sigma_x}^2 \tag{24.2}$$

where ${\sigma_x}^2$ is the sample standard deviation of the x values.

Determining the equation of the line using LINEST()

Excel contains a function to do least squares fitting. Unfortunately if produces a bunch of numbers without indicating what is what. It also has to be configured to do things the way we want. Make sure to configure it to give extra statistics.

*When using **LINEST()** to calculate the least squares fit, always set it to calculate the y-intercept, even when you expect it to be zero! This gives you important information about the data.*

Comparing the result given by the least squares fit using your formulas with your regression output should indicate what several of the quantities are.

(If you use **LINEST()** to do least squares fitting for a lab report, quote the quantities given with the names used in the lab manual. The unidentified block of cells given by Excel is not very meaningful.)

Determining uncertainties in the slope and y-intercept

Case I: Maximum and minimum slopes If the error bars are large enough, then the line of best fit should go through all of the error bars. In this case, there will be two data points which determine coordinates for a line of maximum slope which crosses all of the error bars. Consider the case for positive slope:

If we label two points x_1 and x_2, where $x_1 < x_2$, then we can see from Figure 24.2 that the steepest line which touches the error bars for both x_1 and x_2 is the line between $(x_1 + \Delta x_1, y_1 - \Delta y_1)$ and $(x_2 - \Delta x_2, y_2 + \Delta y_2)$. The slope of this line will then be

$$m_{max} = \frac{(y_2 + \Delta y_2) - (y_1 - \Delta y_1)}{(x_2 - \Delta x_2) - (x_1 + \Delta x_1)}$$

and then the y-intercept is given by

$$b_{min} = (y_1 - \Delta y_1) - m_{max}(x_1 + \Delta x_1) = (y_2 + \Delta y_2) - m_{max}(x_2 - \Delta x_2)$$

Similarly the line with the least slope which touches the error bars for both x_1 and x_2 is the line between $(x_1 - \Delta x_1, y_1 + \Delta y_1)$ and $(x_2 + \Delta x_2, y_2 - \Delta y_2)$. The slope of this line will then be

$$m_{min} = \frac{(y_2 - \Delta y_2) - (y_1 + \Delta y_1)}{(x_2 + \Delta x_2) - (x_1 - \Delta x_1)}$$

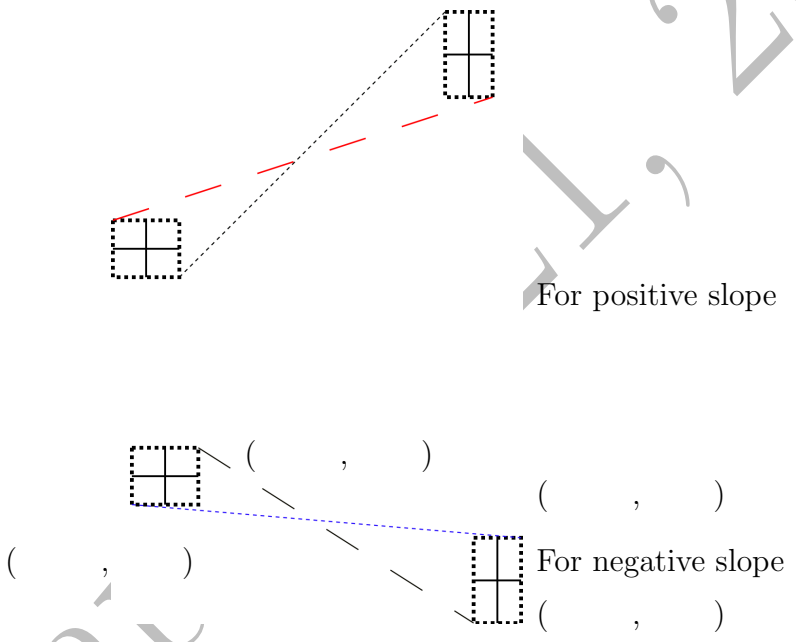


Figure 24.2: Maximum and Minimum Slope Coordinates from a Point

and then the y-intercept is given by

$$b_{max} = (y_1 + \Delta y_1) - m_{min}(x_1 - \Delta x_1) = (y_2 - \Delta y_2) - m_{min}(x_2 + \Delta x_2)$$

The case for a negative slope is shown in Figure 24.2; the analysis is left to the student.

The points for the maximum and minimum slope will not always be the endpoints on the graph.

Index function in Excel To calculate the slope and y-intercept in Excel from a block of data, we can use the *index* function. Its syntax is as follows:

- index(reference, row number, column number)
- reference is the cell range to look in
- row number (starts at one)
- column number (starts at one)

So if we have a data set of 6 values where the x values start in $A2$, and the Δx values start in $D4$, then we can get

$$x_2 + \Delta x_2$$

by the formula

$$= \mathbf{INDEX}(\mathbf{A2} : \mathbf{A7}, \mathbf{2}, \mathbf{1}) + \mathbf{INDEX}(\mathbf{D4} : \mathbf{D9}, \mathbf{2}, \mathbf{1})$$

(Note that the only difference is in which block of data to use.) You'd probably write the formula as

$$= \mathbf{INDEX}(\$\mathbf{A}\$\mathbf{2} : \$\mathbf{A}\$\mathbf{7}, \mathbf{2}, \mathbf{1}) + \mathbf{INDEX}(\$\mathbf{D}\$\mathbf{4} : \$\mathbf{D}\$\mathbf{9}, \mathbf{2}, \mathbf{1})$$

so that you could copy it and still refer to the same blocks of data.

Case II: standard errors If the error bars are small enough, then the points will be scattered in such a way that no line can be drawn which crosses all of the error bars. In this case, the uncertainties in the slope and y-intercept reflect the scatter of the points. In this case, the uncertainty in the slope and y-intercept will be calculated using the **standard errors** in the slope and y-intercept, in much the same way that the uncertainty for an average value is calculated using the *standard error of the mean*.

24.1.3 Displaying Lines

Unless you are going to give the equation of a line or curve, do not *show it on a graph!*

Plotting arbitrary lines

To display a line on the graph, such as a best fit line, one can use a series which has not yet been used. When one knows the *equation* of a line, all one needs is two endpoints so that a line can be drawn between them. To allow this, include 2 values at the end of your x series, x_{min} and x_{max} which are the minimum and maximum values from the x data, respectively. Placing the y values calculated from the line equation in the corresponding cells of another series will allow a line to be plotted between those points. (Set the format for that series to lines only.)

Using "trendline"

There is a built-in feature called *trendline* which allows you to display various fits to data. A *linear* trendline is, in fact, a least squares fit. Unfortunately, this feature does not automatically display the parameters for the fit, so it's not as much use as it could be.

24.2 Recap

By the end of this exercise, you should understand the following terms:

- linear graph
- error bars
- least squares fit
- correlation coefficient
- large scatter of data points
- small scatter of data points

In addition, you should, using Excel, be able to:

- plot a linear graph
- add error bars
- perform a least squares fit
- show the least squares fit line on the graph with the data

You should also be able to

- determine whether the points on a graph classify as either "small" or "large" scatter, and calculate graphical uncertainties appropriately in either case;
- compare different linearizations of the same function and to explain why one may be preferred over others.

Chapter 25

Plotting Curves Using a Spreadsheet

25.1 Theory

Consider the data shown in Figure 25.1.
First of all, the graph does not have error bars. However it it also not smooth. *Dot-to-dot drawings do not usually belong in scientific reports; smooth curves are usually more appropriate.* The following discussion should help you to use a spreadsheet to produce non–linear plots.

25.1.1 Displaying Curves

You have previously seen how a a *line* can be drawn by simply connecting its endpoints. Similarly, a *curve* can be *approximated* by a series of very short line segments between points along the curve. If the points are close enough, the line will look smooth. If we wish to plot a curve in a spreadsheet, proceed as follows:

1. Instead of just using two values, x_{min} and x_{max} , for an x–series, create a series of values $x_i, i = 0 \ldots N$ where

$$x_i = x_{min} + i \left(\frac{x_{max} - x_{min}}{N} \right)$$

   so that

$$x_0 = x_{min}$$

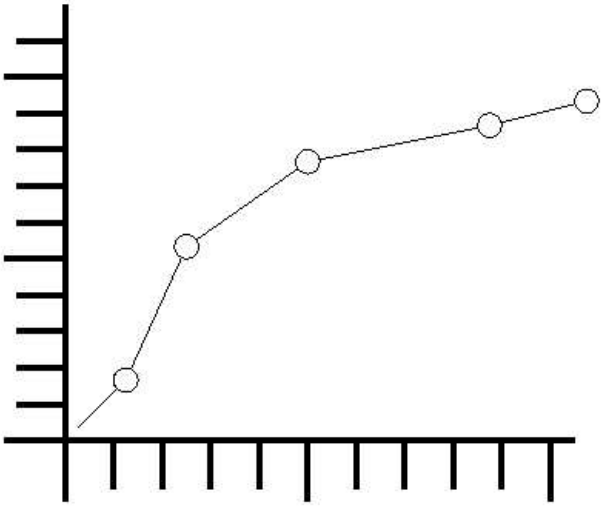


Figure 25.1: Logarithmic Fit the Wrong Way

and

$$x_N = x_{max}$$

and for each point

$$y_i = f(x_i)$$

Note that even though the x values for the first and last point are taken from the original data, the y values are not, since the purpose is to find points on the *curve*.

A value of 100 for N should be sufficient to make a smooth curve. If not, make it bigger. You can try making it smaller if that many points aren't needed.

2. For each of the x_i, calculate the corresponding y_i value from the curve equation and add this series to the graph.

3. Remove the markers and add lines for this series. All points in this series will thereby be joined with line segments. You should be able to produce a graph such as in Figure 25.2. (Error bars have been left off for simplicity. They can be produced in the usual way.)

Note that in the graph of Figure 25.2, none of the data points actually fall on the curve. This is often the case.

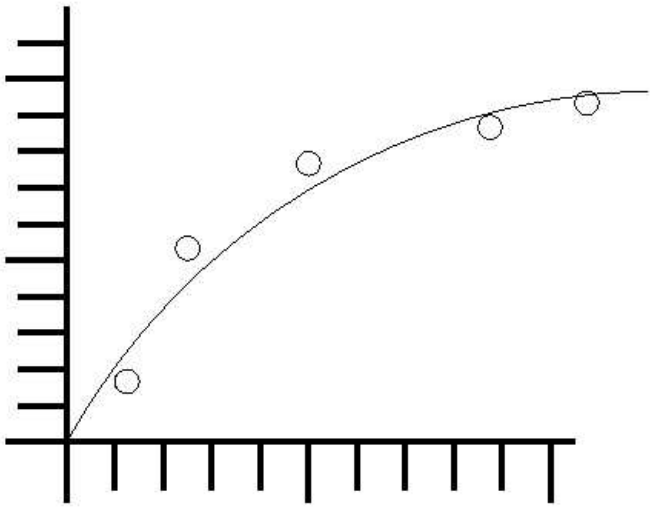


Figure 25.2: Logarithmic Fit the Right Way

25.1.2 Piecewise Defined Functions

In some cases, a curve must be made from data which must be fit to different equations in different regions. In this case the trick comes in trying to join the two fit equations smoothly. In this case,

$$f(x) = \begin{cases} F_1(x), & \text{if } x \leq x_1; \\ F_2(x), & \text{if } x \geq x_2. \end{cases}$$

where $x_2 > x_1$. To make the function $f(x)$ *continuous*, we require that

$$f(x_1) = F_1(x_1)$$

and

$$f(x_2) = F_2(x_2)$$

To produce a *smooth* fit, then in addition we require that

$$f'(x_1) = F_1{}'(x_1)$$

and

$$f'(x_2) = {F_2}'(x_2)$$

These 4 equations can be solved exactly by a polynomial in 4 unknowns, i.e. an equation of the form

$$f(x) = \alpha x^3 + \beta x^2 + \gamma x + \delta$$

and therefore

$$f'(x) = 3\alpha x^2 + 2\beta x + \gamma$$

To solve this, we need to set up a system of equations as follows: If we let

$$\mathcal{A} = \begin{pmatrix} {x_1}^3 & {x_1}^2 & x_1 & 1 \\ {x_2}^3 & {x_2}^2 & x_2 & 1 \\ 3{x_1}^2 & 2x_1 & 1 & 0 \\ 3{x_2}^2 & 2x_2 & 1 & 0 \end{pmatrix}$$

and

$$\mathcal{B} = \begin{pmatrix} F_1(x_1) \\ F_2(x_2) \\ {F_1}'(x_1) \\ {F_2}'(x_2) \end{pmatrix}$$

Then it should be clear that

$$\mathcal{A}X = \mathcal{B}$$

or

$$X = \mathcal{A}^{-1}\mathcal{B}$$

where

$$X = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}$$

(${F_1}'(x_1)$ can be determined by using x_1 and the fit point immediately to its left, and ${F_2}'(x_2)$ can be determined by using x_2 and the fit point immediately to its right.) This system can then be solved using the **matrix invert** and **multiply** features of a spreadsheet.

Note: This will produce a *smooth* graph; whether it produces an *accurate* graph remains to be seen, (although smooth transitions are more common than others). This should produce a graph such as in Figure 25.3. Error bars have not been shown so that the smoothness of the curve can be seen easily. As before, error bars should be included unless you are specifically told to omit them.

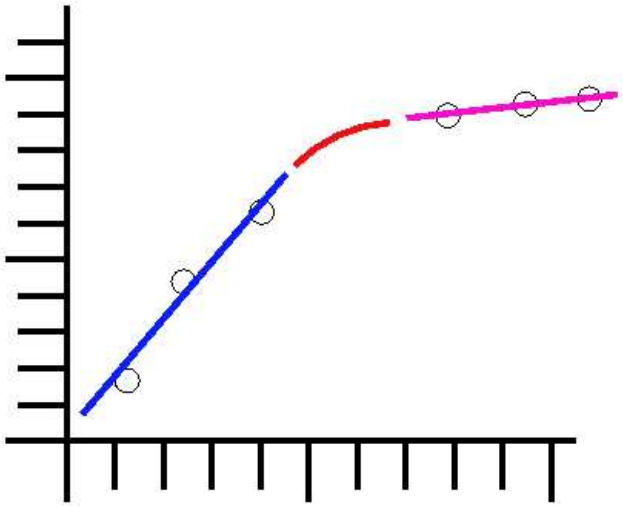


Figure 25.3: Piecewise Defined Function

25.2 Recap

By the end of this exercise, you should understand how to display any curve on a spreadsheet graph.

Chapter 26

Goal Seek Spreadsheet Feature

26.1 Theory

Often when dealing with numerical data, one will be faced with functions which are non–invertible. Assume

$$y = f(x)$$

If $f(x)$ is invertible, then there exists a function $f^{-1}(x)$ such that

$$x = f^{-1}(y)$$

so that given a value y_0, we can calculate the value of x_0 which corresponds to y_0[1] in the original function; ie.

$$y_0 = f(x_0)$$

In many cases, an inverse function does not exist. This can happen for a variety of reasons:

1. The function is many–to–one. Many x values give the same y value, so there is not a *unique* inverse. (eg. $y = \sin(x)$)

2. The function is sufficiently complex that no simple inverse exists. (eg. $y = x + e^x$; $y' > 0$ everywhere, but no simple inverse function exists.)

[1]The value of x for which $f(x) = 0$ is known as a *root* of $f(x)$.

To find the x value which produces a given y value, we must **solve** the equation

$$y_0 = f(x)$$

That is, we must *"find the value of x which makes the equation true"*. Solving this equation is the same as solving the equation

$$f(x) - y_0 = 0$$

Various mathematical algorithms exist for finding good numerical approximations to these solutions, and several spreadsheets, including Excel, have one of these built in.

Although there may be many possible solutions to the above equation, a tool such as the spreadsheet will only produce one at a time. In fact most such tools will only give one solution, and the one given will depend on the **starting value** used. The starting value is a value for x which should be somewhere in the vicinity of the solution being approximated. The caveat is, the more you know about the expected solution(s), the easier it is (they are) to find.

Consider the function shown in Figure 26.1. The desired *solution* is the x–value where the function crosses the x–axis.

One algorithm for solving an equation $F(x) = f(x) - y_0 = 0$ is the **Newton-Rhapson** or **Newton's** method.[2] The algorithm starts with an initial value x_0, the "starting value", and computes a sequence $x_1, x_2, x_3 \ldots x_n$ where x_{n+1} is determined in the following way:

The function $F(x) = 0$ is approximated by its tangent[3] at the point $(x_0, F(x_0))$. The equation of the tangent is

$$F(x_0) + (x - x_0)F'(x_0) = y$$

Then x_1 is the point where the tangent line intersects the x–axis and x_1 is taken as the next approximation to the solution. Thus, for determining x_{n+1}, we get the equation

$$F(x_n) + (x_{n+1} - x_n)F'(x_n) = 0$$

[2]Other methods include the **bisect** method which works if you know one value of x for which $f(x) > 0$ and another for which $f(x) < 0$.

[3]If we can't or don't want to figure out the equation for $F'(x)$, we can approximate it by determining the **secant line** of $F(x)$ at that point. This is the line which joins $(x, F(x))$ and $(x + \delta, F(x + \delta))$, where δ is a very small number.

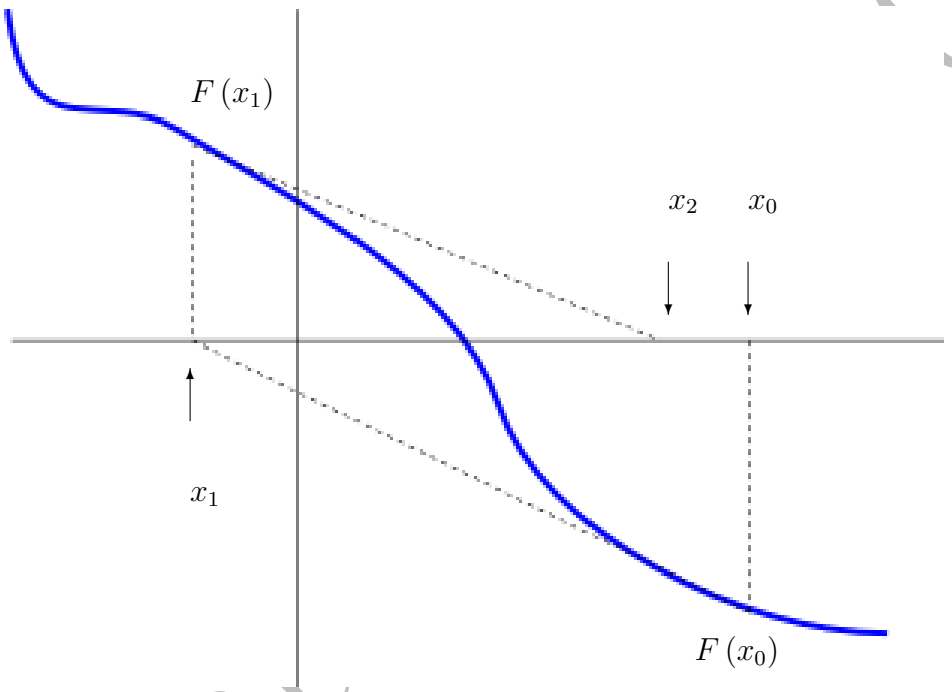


Figure 26.1: Searching Near Root

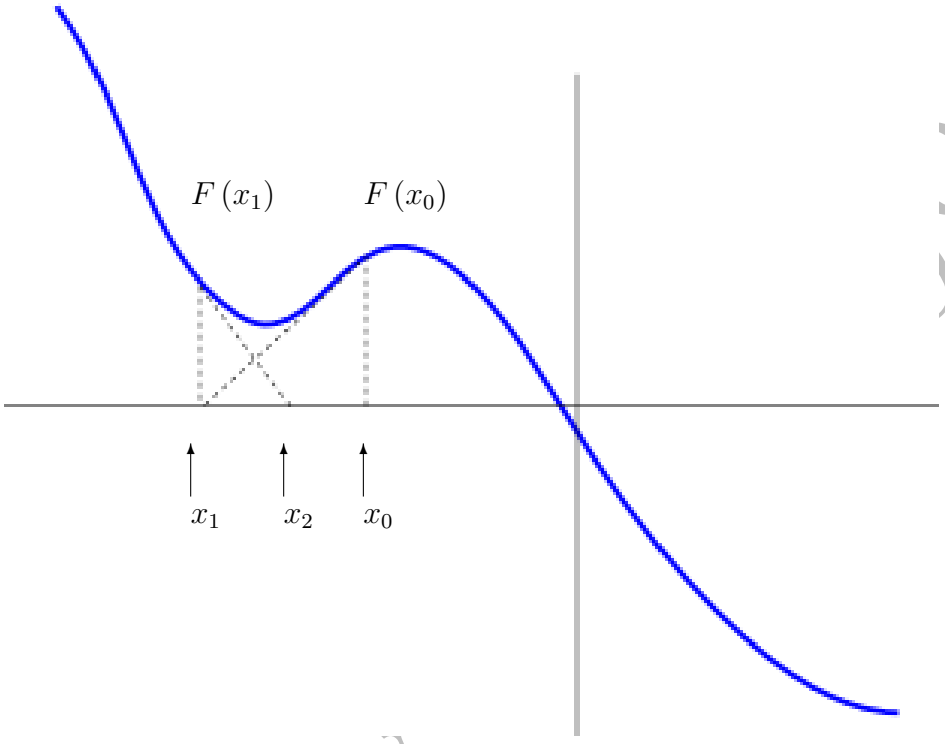


Figure 26.2: Searching Near Local Minimum

Solving for x_{n+1}, we get the equation

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}$$

The algorithm stops when $\left|-\frac{F(x_n)}{F'(x_n)}\right|$ is less than the largest error one is willing to permit in the root.

By starting at point x_0, Newton's method will give us a succession of values which will eventually converge to the solution. However, consider a slightly different function, using the starting point shown in Figure 26.2.

In this case, the algorithm may easily get "stuck" to the left of zero and *never* find the solution, but rather converge to the location of the local minimum. It is even possible for the results to *diverge*, and get farther from the solution at each iteration!

26.1.1 Uncertainties in Parameters

One difficulty with using this method for inverting a function is that we cannot easily generate standard errors in the independent variable. While there is no general way to do this, we can get some idea of the sensitivity of the independent variable to changes in function paramter values by varying the paramter values within their experimental uncertainties. This can be done as follows, for a function with three parameters A, B, and C:

1. Set up new variable values in the spreadsheet so that, for instance,

   $$A' = A + 2(\mathbf{rand()} - 0.5)(\Delta A)$$

   (The same can be done for B and C.) (The **rand()** function of the spreadsheet gives a random value between 0 and 1, so subtracting 0.5 from it gives a random value between -0.5 and $+0.5$, and multiplying this by 2 gives a value between -1 and 1 ; thus we get a value between $A - \Delta A$ and $A + \Delta A$.)

2. Now you will see that any time you recalculate, all of those values (and thus the results) will change as well.

3. Copy the values A', B', and C' and use the *Paste Special* command to paste the values as numbers, *not formulas* into somewhere in the spreadsheet.

4. Change the setup of goalseek to point to the values you just copied instead of the original A, B, and C.

5. To see the effect on the parameters calculated, you will need to repeat the cut-and-paste step and re-run goal seek.

6. Do this 5 or 10 times and use the ranges in the independent variable values as estimates for the uncertainties in the independent value.

(You might wonder why we don't just change all of the values to their values with uncertainties added, and then see how much that changes the answer, thus giving us the maximum uncertainty. The problem is that in complicated calculations, the effects may cancel each other out, so there is no apparent uncertainty. If we were subtracting the numbers, this is obvious.)

26.1.2 Example: Non-Invertible Functions

Suppose we have a function

$$y = Ax^2 + B/x$$

and we want to find the value of x for which $y = 5$. If we want to find the value of x such that

$$5 = Ax^2 + B/x$$

then that is the same as finding the value of x for which

$$5 - \left(Ax^2 + B/x\right) = 0$$

In this case, A and B are constants, and we wish to find the value(s) of x for which this is true, so A and B are our "data", and x is our "result".

1. Set up a cell containing the formula

   $$5 - \left(Ax^2 + B/x\right) = 0$$

   where A, B, and x are references to variable cells. The first two are parameters in your equation, and the third is your "starting value" for x. Provide values for all of these cells.

2. Set up goal seek with x as your variable, and the cell containing $5 - (Ax^2 + B/x)$ as your formula cell. Make your target value 0.

3. Run goal seek a few times to see if your results converge. If your results do not converge, try other starting values.[4]

To get some idea of the sensitivity of the parameters to changes in data values by varying the data within its experimental uncertainty, proceed as follows. (Note in this case the variables are the parameters A and B.)

1. Create new data values above so that, for instance,

   $$A' = A + 2(\mathbf{rand}() - 0.5)(\Delta A)$$

   and

   $$B' = B + 2(\mathbf{rand}() - 0.5)(\Delta B)$$

[4] If you know enough about the function to be sure that it has a root, then there should be some starting value which will work. You just have to find it.

2. Now you will see that any time you recalculate, those two values will change as well. Copy these A' and B' values and use the *Paste Special* command to paste the values as numbers, *not formulas* into another part of the spreadsheet.

3. Change your formula cell to point to these new copied A' and B' values instead of the previous A and B. Now the the solution, x will depend on the uncertainties in A and B.

If you don't point to the pasted values, then the random variation will be happening *while* goalseek is running, which could make your results very unpredictable.

4. To see the effect of the uncertainties on the parameters calculated, you will need to repeat step 2 above and then re-run goal seek.

5. Do this 5 or 10 times and use the range in the solution as an estimate for the uncertainty.

26.1.3 Goal Seek in Excel ™

Goalseek is a tool in Excel, which is simple to use. When you invoke it from the menu, you will be presented with a dialogue box as shown in Figure 26.3.

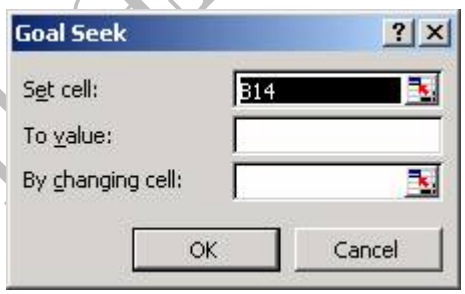


Figure 26.3: Goal Seek parameters

Set cell Select the cell for the modified function.

To value 0

By Changing Cell Select the cell for the independent variable.

26.2 Recap

By the end of this exercise, you should know how to use the *goalseek* function to work with non-invertible functions.

Chapter 27

Solver Spreadsheet Feature

27.1 Theory

At times you will wish to find a maximum or minimum value for a function over a given range of input values. For instance, given

$$y = f(x)$$

one might need to know what value of x maximizes the value of y. If $f(x)$ is monotonically increasing or decreasing, then the function will have its maximum value at one of the endpoints of the interval in question. In other cases, it may be much harder to determine. If the function is continuous, then the function will have a local maximum or minimum wherever

$$y' = 0$$

These points can be approximated in a similar way to the way roots of a function are found, as just described. In this case, if we use successive values of the function to estimate both the first AND second derivatives of the function at a point, then it is possible to estimate the point at which the slope will be zero.

There are many numerical algorithms for finding good numerical approximations to these solutions, and Excel Pro has one of these built in.

Although there may be many possible candidates for the solution to the above equation, the spreadsheet will only check one such candidate, and the one checked will depend on the **starting value** given. The starting value is a value for x which should be somewhere in the vicinity of the candidate

being approximated. The caveat is, the more you know about the expected solution(s), the easier it is (they are) to find. (In practice, most functions are sufficiently well-behaved that there may be at most a handful of possible candidates, and usually only one. In these cases, such algorithms will usually do a good job of approximating them.)

To find the x value which maximizes y value, we must **optimize** the equation

$$y = f(x)$$

That is, we must *"find the value of x which gives the maximum (or minimum) value for y"*.

A least squares fit is really just a kind of optimization; it amounts to finding the values of m and b (for a linear fit) which minimize the sum of squares error.

In this case, the function to be minimized is

$$S = S(m, b)$$

This is a function of 2 variables, and we want to find the values of both m and b which minimize S.

For a non-linear fit, then the equation for S and thus the parameters involved would be different.

27.1.1 Uncertainties in Parameters

One difficulty with using these methods for fitting is that we cannot easily generate standard errors in the fit parameters. While there is no general way to do this, we can get some idea of the sensitivity of the parameters to changes in data values by varying the data values within their experimental uncertainties. This can be done for a function with parameters A and B as follows:

1. Replace the variable values in the data so that, for instance,

$$x_i' = x_i + 2(\mathbf{rand}() - 0.5)(\Delta x_i)$$

where the x_i are the given values with uncertainties in them. (The same can be done for y values.) (The **rand()** function of the spreadsheet gives a random value between 0 and 1, so subtracting 0.5 from it gives a random value between -0.5 and $+0.5$, and multiplying this by 2 gives

a value between -1 and 1 ; thus we get a value between $x_i - \Delta x_i$ and $x_i + \Delta x_i$.)

2. Now you will see that any time you recalculate, all of those values (and thus the results) will change as well.

3. Copy these x_i' and x_i' values as *values* to another part of the spreadsheet.

4. Change your formula cell to point to these new copied x_i' and x_i' values instead of the previous x_i and y_i. Now the the solution A and B values will depend on the uncertainties in x and y.

5. To see the effect on the parameters calculated, you will need to re-run the solver.

6. Do this 5 or 10 times and use the ranges in the parameter values as estimates for the uncertainties in the parameters.

(You might wonder why we don't just change all of the data values to their values with uncertainties added, and then see how much that changes the answer, thus giving us the maximum uncertainty. The problem is that in complicated calculations, the effects may cancel each other out, so there is no apparent uncertainty. If we were subtracting the numbers, this is obvious.)

27.1.2 Solver Example: Least Squares Fitting

Getting the parameter values

Consider the data in Table 27.1.

In this case, we want to do a non-linear fit. Specifically, we wish to minimize the value of S where

$$S = S(A, B)$$

where

$$S = \sum (\hat{y}_i - y_i)^2$$

At first glance, S does not seem to depend on A and B. However, if we expand the above equation we get

$$S = \sum \left(A{x_i}^2 + B - y_i\right)^2$$

i	x_i	y_i	$\hat{y}_i = Ax_i{}^2 + B$	$(\hat{y}_i - y_i)^2$
1	0.40	0.00	0.28	0.08
2	0.77	2.00	1.46	0.29
3	1.35	2.70	3.32	0.38
4	1.72	4.30	4.50	0.04
S= $\sum (\hat{y}_i - y_i)^2$				0.79

Table 27.1: Least Squares Fit data

since

$$\hat{y}_i = Ax_i{}^2 + B$$

Note that if we change A and /or B, then the values of $\hat{y}$ will change and thus S will change also.

If we set up a spreadsheet like Table 27.1, then one can vary A and B manually and observe the changes in S. Having done so, it is easy to automate the process using the solver in the spreadsheet.

Note also that we could have used any function in place of the equation $y = Ax^2 + B$ in the fourth column above, and the calculation of S would remain the same. Thus it now becomes easy to do least squares fits of *non-linear* functions as well as linear ones.

1. Set up a cell containing the formula

   $$\sum (\hat{y}_i - y_i)^2$$

   where $\hat{y}_i$, and y_i, are references to variable cells.

2. Set up the solver with A and B as your variables, and the cell containing $\sum (\hat{y}_i - y_i)^2$ as your solution cell. Configure the solver to minimize this.

3. Run the solver a few times to see if your results converge.

27.1.3 Getting the parameter uncertainties

To get some idea of the sensitivity of the parameters to changes in data values by varying the data within its experimental uncertainty, proceed as follows. (Note in this case the variables are the actual data point coordinates themselves.)

i	original x_i	Δx_i	original y_i	Δy_i	x_i'	y_i'
1	0.40	0.01	0.00	0.05	0.39	0.04
2	0.77	0.02	2.00	0.05	0.76	2.03
3	1.35	0.03	2.70	0.05	1.36	2.71
4	1.72	0.05	4.30	0.05	1.7	4.34

Table 27.2: Finding parameter sensitivities

1. Make new columns of data so that, for instance,

$$x_i' = x_i + 2(\mathbf{rand}() - 0.5)(\Delta x_i)$$

and

$$y_i' = y_i + 2(\mathbf{rand}() - 0.5)(\Delta y_i)$$

as shown in Table 27.2.

2. Now you will see that any time you recalculate, all of those values will change as well. Copy these x_i' and y_i' values and use the *Paste Special* command to paste the values as numbers, *not formulas* into somewhere in the spreadsheet.

3. Change your formulae for $(\hat{y}_i - y_i)^2$ so that it is these new values for y_i' are used instead of the y_i values from before.

If you don't point to the pasted values, then the random variation will be happening *while* the solver is running, which could make your results very unpredictable.

4. Run the solver. Now the SSE will depend on the uncertainties in x and y.

5. To see the effect of the uncertainties on the parameters calculated, you will need to repeat the cut-and-paste and then re-run the solver.

6. Do this 5 or 10 times and use the ranges in the parameter values as estimates for the uncertainties in the parameters.

27.1.4 Solver in Excel ™

Solver is an add-in tool in Excel, which is simple to use. To be able to use it, do the following (in Excel 2007):

1. Click on the "Orb"; (i.e. the office icon)
2. Click on the "Excel options" button at the bottom of the dialog.
3. Select "add-ins".
4. Select "Solver add-in".

It will then show up under "Analysis" on the "Data" tab.

When you invoke it, you will be presented with a dialogue box as shown in Figure 27.1.

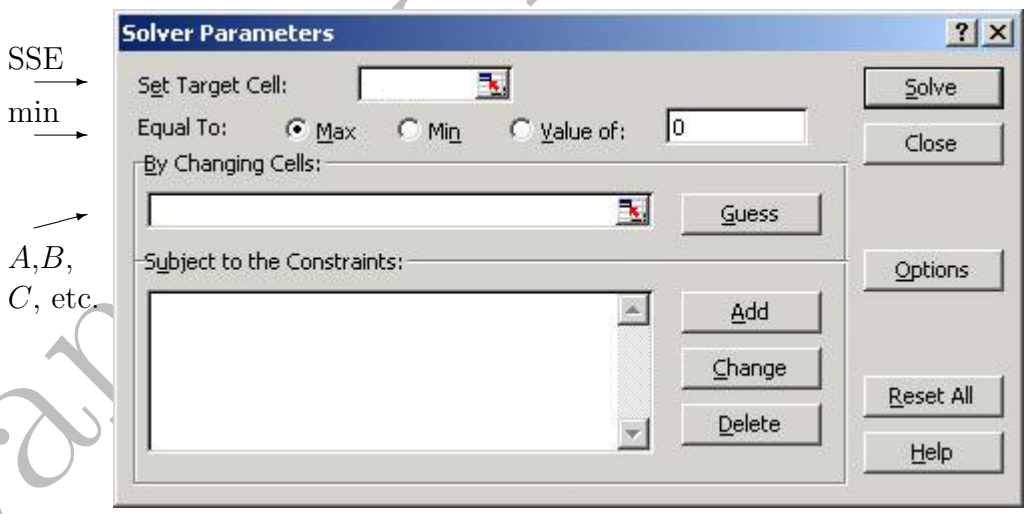


Figure 27.1: Solver parameters

Set Target Cell In our example of a least squares fit, this would be the SSE

Equal To There are three choices here:

- min
- max
- value

In the case of the least squares fit, we would like to minimize the SSE.

By Changing Cells All of the parameters of our fit need to be changed in order to find the solution.

Buttons

Solve

Options There are some options you can choose, which are indicated by a dialogue box as shown in Figure 27.2.

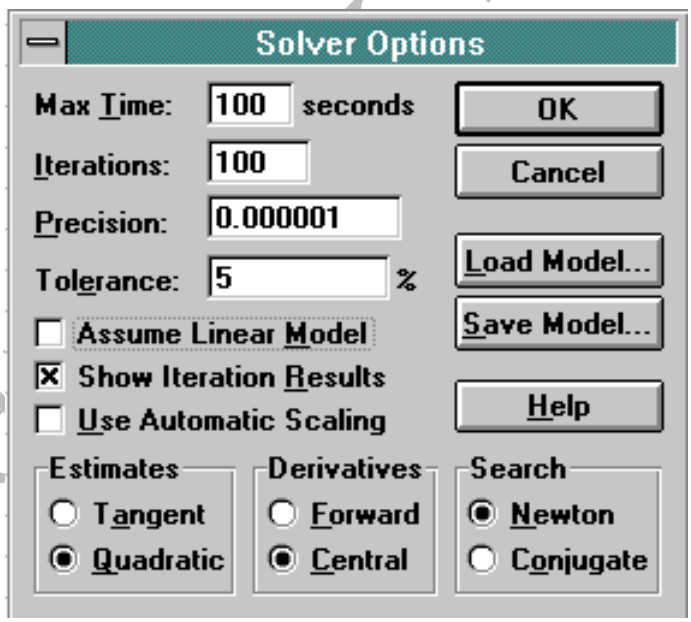


Figure 27.2: Solver options

Max Time

Iterations

Precision

Tolerance (%)

Buttons

OK There are other options about the type of algorithm used; usually the default is fine. You can try others if you wish.

When the solver is finished, you should see a dialogue box as shown in Figure 27.3.

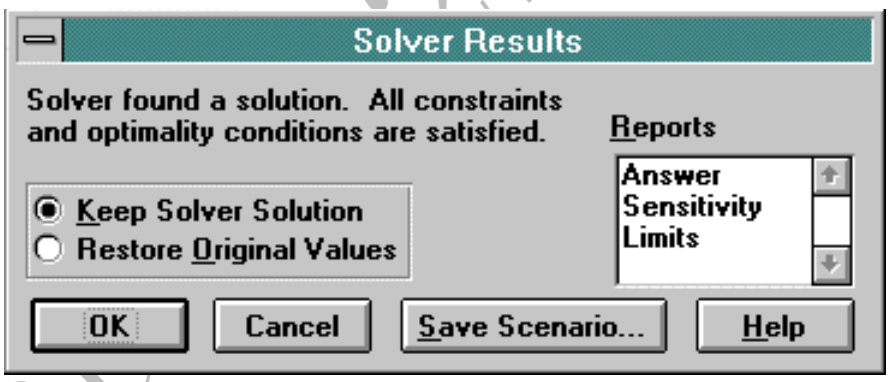


Figure 27.3: Solver results

Buttons

Keep Solver Solution

Restore Original Values

OK

27.2 Recap

By the end of this exercise, you should know how to use the *solver* function to optimize functions.

Chapter 28

Sketching Functions and Extracting Parameters

28.1 Introduction

Previously you learned how to linearize functions and then use least squares fitting to get parameters from the graph. Then you learned how to use spreadsheet features to do least squares fitting with non-linearized functions. However, to do that it is necessary to have reasonable starting values for the parameters involved. Now you are going to learn how to do that.

In first year calculus, you should have learned about sketching functions. This will be discussed now because of the way that parameters for functions can be extracted from plots of observed data; in other words, knowing how certain parameters affect the graph of a function allows one to determine those parameters from the graph of a function.

28.1.1 Critical Points

A critical point of a function occurs where the slope is zero. In other words,

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = 0$$

If a graph of actual data has a critical point, then this can be used according to the equation above. (Critical points also occur where the function is undefined. These may be observable as well.)

28.1.2 Limits and Asymptotes

The limits of a function as $x \to \infty$ or any other point will often be useful if either of the following is true:

- function has a limit; in this case, if we can observe values of the function near enough to the limit, we can approximate the limit. Keep in mind that for one function, there may be useful limits as you approach a variety of x values, so you may need to do this several times to extract as much information as possible.

- function has an asymptote; an asymptote is just a special kind of limit. In this case, for a function $f(x)$, an asymptote exists if

$$\lim_{x\to\infty} f(x) - (mx + b) = 0$$

for some choice of m and b. If they exist, then m and b can be found as follows:

1. Calculate m by using the fact that

$$\lim_{x\to\infty} \frac{\mathrm{d}f(x)}{\mathrm{d}x} = m$$

ie. See if you can observe the data approaching an asymptote. If so, then you can draw in the line and calculate the slope m.

2. Calculate b according to the equation above, using this value of m. In other words,

$$\lim_{x\to\infty} f(x) - (mx + b) = 0$$

thus

$$\lim_{x\to\infty} f(x) - mx - b = 0$$

and so

$$\lim_{x\to\infty} f(x) - mx = b$$

ie. extrapolate the asymptote you have drawn to where it crosses the y axis and determine b. (You don't actually need to do this; given any two points on a line you can calculate both m and b.)

Thus if asymptotes exist, they can be used to determine parameters of the original function.

28.1.3 Roots

If a function has *zeroes*, i.e. points where $f(x) = 0$, then these can also be useful if they can be observed in the data. (For that matter, if the function takes on any fixed value then this can be used.)

28.1.4 Modified Functions

At times, some of the above rules will be useful if applied to a modified version of the function; for instance $f(x)$ may not have an asymptote, but $\ln(f(x))$ might. With practice you should get better at spotting these situations.

28.1.5 Substitution

If all but one of the parameters for a function can be estimated by using the previous methods, then the final parameter can be estimated by substituting the other values into the equation for a single data point.

Appendix A

Information about Measuring Instruments

Fill this in as you use new measuring instruments so you will have a reliable reference. Put frequently used instruments in Table A.1 and experiment-specific ones in Table A.2.

ref. #	measuring instrument	precision measure	range	units
A1	vernier caliper			
A2	micrometer caliper			
A3	stopwatch			
A4	GL100R balance			
A5				
A6				

Table A.1: Measuring instrument information

ref. #	measuring instrument	precision measure	range	units
B1	spring scale A			
B2	spring scale B			
B3				
B4				
B5				
B6				
B7				
B8				
B9				
B10				
B11				
B12				
B13				
B14				
B15				

Table A.2: Measuring instrument information (continued)

Appendix B

Common Uncertainty Results

Following are some common results about uncertainties which you may find useful. If there are others which you feel should be here, inform the lab supervisor so that they may included in future versions of the lab manual.

$$\Delta(\overline{x}) = \overline{(\Delta x)}$$

$$\Delta(x^n) \approx n|x|^{n-1}\Delta x$$

$$\Delta(\sin x_R) \approx |\cos x_R|\,(\Delta x)_R$$

$$\Delta(\tan x_R) \approx (\sec x_R)^2(\Delta x)_R$$

where x_R denotes x in *radians*.

$$\Delta \ln x \approx \frac{1}{x}\Delta x = \frac{\Delta x}{x}$$

$$\Delta x^y \approx \left|x^{y-1}y\right|\Delta x + \left|x^y \quad \ln x\right|\Delta y$$

$$\Delta \sqrt[y]{x} \approx \left|x^{\left(\frac{1}{y}-1\right)} \quad \frac{1}{y}\right|\Delta x + \left|x^{\frac{1}{y}} \quad \ln x\right|\frac{\Delta y}{y^2}$$

$$\Delta f(x,y,z) \approx \left|\frac{\partial f}{\partial x}\right|\Delta x + \left|\frac{\partial f}{\partial y}\right|\Delta y + \left|\frac{\partial f}{\partial z}\right|\Delta z$$

Appendix C

Common Approximations

Following are some common approximations which you may find useful. If there are others which you feel should be here, inform the lab supervisor so that they may included in future versions of the lab manual.

Taylor Series Expansion

$$f(x+h) = \sum \frac{f^n(x)}{n!} \approx f(x) + hf'(x)$$

The following derive from the Taylor series expansions, where $x \ll 1$. In cases where an approximation is given with more than one term, the first term alone may be sufficient in some cases.

$$(1+x^n) \approx 1 + nx$$

$$\ln(1+x) \approx x - \frac{x^2}{2}$$

$$e^x \approx 1 + x + \frac{x^2}{2!}$$

The following also assume x is in *radians*.

$$\sin(x) \approx x - \frac{x^3}{3!}$$

$$\cos(x) \approx 1 - \frac{x^2}{2!}$$

Appendix D

Use of the Standard Form for Numbers

D.1 Introduction

Learning how to use the standard form is easy; learning *when* to use it is a bit harder. Following are some examples which should help you decide when to use it. It is perhaps easiest to illustrate by comparing results using the standard form and not using it. Remember the point is to present things more concisely.

1. Voltage V of 2.941 Volts; uncertainty of 0.4517 Volts

   - First round the uncertainty to one significant figure; thus

     $\Delta V \approx 0.4$ V
   - Next round the quantity so the last digit displayed is in the same decimal place as the uncertainty; thus

     $V \approx 2.9$ V (when rounded to the same decimal place as the uncertainty)
   - So

     $V = 2.9 \pm 0.4$ V

     In this case, scientific notation is not needed.

2. Mass m of 140.6 grams; uncertainty of 531.7 grams

   This is an example of a situation where the uncertainty is larger than the quantity itself. The process involved is the same.

   - First round the uncertainty to one significant figure; thus

     $\Delta m \approx 500$ g
   - Next round the quantity so the last digit displayed is in the same decimal place as the uncertainty; thus

     $m \approx 100$ g (when rounded to the hundreds place)
   - So

     $m = 100 \pm 500$ g

     In this case, scientific notation *is* needed, because we need to get rid of the placeholder zeroes. One option would be to write

     $m = (1 \pm 5) \times 10^2$ g

     Another option would be to write

     $m = 0.1 \pm 0.5$ kg

     Note that this last option is more concise.

3. Diameter d of 0.727 cm; uncertainty of 0.015 cm

   - First round the uncertainty to one significant figure; thus

     $\Delta d \approx 0.02$ cm
   - Next round the quantity so the last digit displayed is in the same decimal place as the uncertainty; thus

     $d \approx 0.73$ cm (when rounded to the hundredths place)
   - So

     $d = 0.73 \pm 0.02$ cm

     In this case, scientific notation is not needed, because there are no placeholder zeroes.

     Another option would be to write

     $d = 7.3 \pm 0.2$ mm

Note that this last option is slightly more concise, since it gets rid of two zeroes, and puts d in proper form for scientific notation (even though in this case the exponent would be zero).

4. Time t of 943 s; uncertainty of 29 s

   - First round the uncertainty to one significant figure; thus

     $\Delta t \approx 30$ s
   - Next round the quantity so the last digit displayed is in the same decimal place as the uncertainty; thus

     $t \approx 940$ s (when rounded to the tens place)
   - So

     $t = 940 \pm 30$ s

     In this case, scientific notation is needed, because there are placeholder zeroes.

     One option would be to write

     $t = (9.4 \pm 0.3) \times 10^2$ s

     This is in the standard form.

     Another option would be to write

     $t = 9.4 \pm 0.3$ hs

     However, since "hectoseconds" are not commonly used I would avoid this (although it is also correct).

     Similarly you could write

     $t = 94 \pm 3$ das

     Again, since "dekaseconds" are not commonly used I would avoid this (although it is also correct).[1]

[1] I actually had to check on the spelling and notation for dekaseconds and hectoseconds, so that illustrates how (un)familiar they are.

Appendix E

Order of Magnitude Calculations

To check your conversions, (among other things), do a rough calculation of your results carrying every value to just 1 significant[1] figure. (This is easy to do quickly on a piece of paper without a calculator, so you can be sure that calculated results are in the right ballpark.)
For instance, for the *"Measuring 'g'"* experiment,

$$h \approx 5m$$

$$t \approx 1s$$

Thus

$$g \approx \frac{2 \times 5}{1^2} \approx 10m/s^2$$

Be sure to write out these calculations; that way you'll be clear on the units you used, etc. If you make an error and have to correct it, you'll want a record of it so you don't make it when you do the "real" calculations.
Since this result is about what you'd expect, then you know any values in that range should be reasonable.

E.1 Why use just one or two digits?

There are a couple of reasons:

[1] If the digit is a 1 or a two, then you may carry 2 figures. If you do this then your answer should be within about 10% of the value you'd get with a detailed calculation. This is easily close enough to spot any major errors.

1. Since we're only using one data point, instead of all of our values, the result will be approximate, so the extra digits aren't needed.

2. When people use calculators, they tend to just automatically write down any answer without thinking. If they made a typing mistake, they often don't notice. So by doing it by hand using only one or two digits, they keep their brains engaged and are more likely to notice an error.

E.2 Using the median instead of the average

If you have several measurements of a quantity, do the calculation with one value instead of averaging all of them. The median is easy to find, and should be close to the average.

E.3 Order of Magnitude Calculations for Uncertainties

In a similar way, you can check to see if your uncertainties are reasonable. In the above example, if

$$\Delta h \approx 5 \text{ cm}$$

and

$$\Delta t \approx 0.1 \text{ s}$$

then

$$\begin{aligned} \Delta g &\approx g\left(\frac{\Delta h}{h} + 2\frac{\Delta t}{t}\right) \\ &\approx 10\left(\frac{0.05}{5} + 2\frac{0.1}{1}\right) \\ &\approx 10\,(0.01 + 0.2) \\ &\approx 10\,(0.21) \\ &\approx 0.2 \text{ m/s}^2 \end{aligned}$$

This makes sense, and so your detailed uncertainty calculations should produce something in this ballpark.

Appendix F

Derivation of the Least Squares Fit

Following is a simple derivation of the least squares fit.

Suppose the relationship between the two experimental parameters being studied is

$$y = f(x)$$

where x is the independent parameter which is varied, and y is the dependent parameter. If $f(x)$ is a polynomial function, or can be approximated by a polynomial, then the least squares method is a *linear* one, and it will almost always give reliable answers. If $f(x)$ cannot be expressed as a polynomial, but consists of transcendental functions, the least squares method is non-linear, and may or may not work reliably. In some cases, a change of variables may result in a polynomial, as in the exponential example above. A function like

$$y = a + \frac{b}{x} + \frac{c}{x^2}$$

is not a polynomial in x, but it is a polynomial in the variable $z = 1/x$.

Suppose the functional relationship between x and y is a polynomial of degree ℓ:

$$y = a_0 + a_1 x + a_2 x^2 \ldots a_\ell x^\ell \tag{F.1}$$

or

$$y = \sum_{j=0}^{\ell} a_j x^j \tag{F.2}$$

and we have a set of N data points x_i, y_i obtained by experiment. The goal is to find the values of the $\ell + 1$ parameters $a_0, a_1 \ldots a_\ell$ which will give the best fit of Equation F.1 to our data points. The first piece of information to note is that

$$N \geq \ell + 1 \tag{F.3}$$

or else we will not be able to make a unique determination. For example, if $\ell = 1$, we need at least two data points to find the equation of the straight line. In order to make any meaningful statistical statements, however, we will need even more than $\ell + 1$ points, as we shall see later. A good rule of thumb: if we wish to fit our data with a polynomial of degree ℓ in a 95% confidence interval, we should choose N such that

$$N - (\ell + 1) \geq 10 \tag{F.4}$$

The idea behind the linear least squares method is to *minimize* the sum

$$S = \sum_{i=1}^{N} \left(y_i - \sum_{j=0}^{\ell} a_j x_i^j \right)^2 \tag{F.5}$$

S will be a minimum if

$$\frac{\partial S}{\partial a_k} = 0 \qquad k = 0, 1, 2 \ldots \ell \tag{F.6}$$

The result will be $\ell + 1$ linear equations in $\ell + 1$ unknowns:

$$\sum_{j=0}^{\ell} a_j \left(\sum_{i=1}^{N} x_i^{j+k} \right) = \sum_{i=1}^{N} x_i^k y_i \qquad k = 0, 1 \ldots \ell \tag{F.7}$$

which can be solved by standard matrix techniques for the unknown coefficients $a_0, a_1 \ldots a_\ell$. As an example, let us consider the case where $\ell = 1$, or

$$y = mx + b$$

In this case,

$$S = \sum_{i=1}^{N} (y_i - (mx_i + b))^2$$

Expanding Equation F.7, we have

$$b(N) + m\left(\sum_{i=1}^{N} x_i\right) = \sum_{i=1}^{N} y_i \tag{F.8}$$

$$b\left(\sum_{i=1}^{N} x_i\right) + m\left(\sum_{i=1}^{N} x_i^2\right) = \sum_{i=1}^{N} x_i y_i \tag{F.9}$$

Then the intercept b and the slope m can be found from Cramer's rule

$$b = \frac{\left(\sum y_i\right)\left(\sum x_i^2\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \tag{F.10}$$

and

$$m = \frac{N\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \tag{F.11}$$

Appendix G

More Advanced Uncertainty (Error) Analysis

Physics is said to be an exact science, but only within a restricted definition of the word exact. It is exact to the extent that an observable *system* can be described in terms of a mathematical model — a set of relationships between the measurable parameters of the system and its properties and/or behaviour. Ultimately, the goal of physics is to find the systems which are capable of being described by mathematical models, formulating those models, and deducing all of the observable consequences of each. If a particular model of a physical system is to be accepted as a valid description of that system, it must be brought to the judgment seat of *experiment*, and that is where the term 'exact science' cannot be rigorously applied. For although calculations based on a particular mathematical model can be carried out, in principle, exactly (i.e., to any number of significant figures), the results of experimental measurements can *never* be called exact. For example, imagine attempting to measure the length of a steel bar with a meter stick. At best, the length can be determined *reliably* to no better than about 0.5 mm, and it would be ludicrous to imagine that its length could be determined to the nearest micron (10^{-6} meter) with such a device. Thus, the length of the bar could never be known exactly (to an infinite number of decimal places). There will always be an *uncertainty* in the length of the bar due to the inherent limitation in the precision of the meter stick.

A mathematical model must be tested by deducing the consequences of setting up the system it represents in a well defined manner, and then observing its behaviour. Basically, the model will state:

1. Set up the system with a given set of initial conditions and constraints.
2. Measure the values of such-and-such parameters.
3. Use those values in the equations describing the model, and perform such-and-such calculations, and compare the experimental results with the model predictions.

Because there are uncertainties in the set-up and measurements involved in the experiment, the physicist can only approximate the required conditions and the values of each parameter. If the experiment is to have any meaning at all, the effects of these uncertainties on the final results must be determined. First, the *causes* of possible uncertainties must be examined and accounted for — this implies a thorough understanding of both the physics of the system under study and the physics of the measurement processes used. Second, the physicist must have the mathematical tools to analyze these things properly. When the causes and effects of **experimental errors** have been examined, the careful experimenter will be able to *eliminate* (or correct for) some of them, and to *design the experimental method in such a way as to minimize the effects of* the remaining ones. With this background in mind, let us turn to a quantitative discussion of experimental errors.

G.1 Estimation of the Mean and Standard Deviation

If a quantity is measured several times, it is usually desirable to end up with *one* characteristic value for the quantity. (Quoting all the data is more precise, but often the original question behind the experiment will not be adequately answered by the raw data).

There are 3 common values which are extracted from data distributions which may be considered "characteristic" in certain circumstances. They are the **mean**, the **median**, and the **mode**. The *mean* is simply the arithmetic average, with which you are familiar. The *median* is the value for which half of the measurements are greater and half of the measurements are less. The *mode* is the most commonly occurring value. Depending on the *reason* for the experiment, the choice of a characteristic answer may change.

In a Gaussian, or **normal** distribution, the above decision is simplified by the fact that the mean, median, and mode all have the same value. Thus, *if*

the data is expected to fit such a distribution, then an average will probably be a good choice as a quantity characteristic of all of the measurements. The uncertainty in this characteristic number will be a reflection of the distribution of the data. Since the variations in the observations are governed by chance, one may apply the laws of statistics to them and arrive at certain definite conclusions about the magnitude of the uncertainties. No attempt will be made to derive these laws but the ones we need will simply be stated in the following sections.

For random errors which follow a normal distribution the mean and standard deviation are determined from an *infinite* set of measurements. This infinite set is known as the *population*. The measurement process consists of obtaining a finite number of *samples* from the population. Consequently, one can only *estimate* the mean and standard deviation of the entire population from a limited sample (the set of n measurements). Suppose you have a sample of n measurements x_i. The best estimate of μ_0, the *population* mean is the *sample* mean, as shown in Equation G.1.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{G.1}$$

The best estimate of σ_0 is given by Equation G.2.

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\frac{(\sum_{i=1}^{n} x_i^2) - n\overline{x}^2}{n-1}} \tag{G.2}$$

Recall that the standard deviation σ is the root mean square deviation of a *single* measurement. The sample mean, however, was determined from n measurements. Hence, $\overline{x}$ must be a better estimate of the population mean than any single measurement, and the *standard error* of the mean or sometimes the *standard deviation of the mean* is given by Equation G.3.

$$\alpha = \frac{\sigma}{\sqrt{n}} \tag{G.3}$$

The standard deviation also has a standard error associated with it, given by Equation G.4.

$$\alpha_\sigma = \frac{\sigma}{\sqrt{2n}} \tag{G.4}$$

Thus, when reporting the result of a series of measurements, in first year you would write $\overline{x} \pm \alpha(n \text{ measurements})$.

G.2 Confidence Interval for A Single Set of Data

The above discussion of the standard deviation of the mean is not quite complete. Specifically, using α as the uncertainty in the calculation of a mean becomes questionable if the amount of data involved is very small (usually less than 30 data values is considered small). To correct for this situation, we can calculate another quantity called a **confidence interval** which is based on the standard deviation of the mean but corrects for the effects of a small number of measurements. First, one must decide what *confidence interval* is required, i.e., *the desired probability that the population mean lies within 1 confidence interval of the sample mean.* Typically, a 95% confidence interval is used. This means that if errors are random, one can assume that the calculated sample mean will differ from the population mean by less than this amount 95% of the time. To calculate this quantity, one proceeds as follows:

From the mean and standard deviation, as calculated above, one then one computes the value of Equation G.5.

$$\Delta x = t_{p,n-1}\frac{\sigma}{\sqrt{n}} = t_{p,n-1}\alpha \tag{G.5}$$

where P is the confidence interval percentage, (typically 95),

$$p = \frac{1}{2}\left(1 + \frac{P}{100}\right) \tag{G.6}$$

and $n-1$ is the number of **degrees of freedom**. Using the Student's t distribution, the value of $t_{p,n-1}$ is found, and Δx can be determined. (For a 95% confidence interval, $p = 0.975$.) When Δx is less than the instrumental precision, enough measurements have been obtained, and the standard error in $\overline{x}$ is just the instrumental precision. Values of $t_{p,n-1}$ are listed in the T–Distribution Table at the end of this section for various confidence intervals and degrees of freedom.

G.3 Confidence Interval for Data Fitted to a Straight Line

When fitting data to a line, we may use the technique of *least squares fitting*. If we wish to fit a straight line to data in the form

$$y = mx + b$$

then the least squares fit gives values for b and m as in Equation G.7 and Equation G.8.

$$b = \frac{\left(\sum y_i\right)\left(\sum x_i^2\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \tag{G.7}$$

$$m = \frac{n\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} \tag{G.8}$$

Once these values for the slope and intercept are determined, the sum of squares error, S is computed. For the linear case, S can be calculated as in Equation G.9.

$$S = \sum y_i^2 - m\sum x_i y_i - b\sum y_i \tag{G.9}$$

In order to estimate the uncertainty in each parameter, the standard deviation σ is computed in Equation G.10.

$$\sigma = \sqrt{\frac{S}{n-2}} \tag{G.10}$$

The **standard error** in the intercept is computed in Equation G.11.

$$\alpha_b = \sigma\sqrt{\frac{\sum x_i^2}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} \tag{G.11}$$

and the standard error in the slope is computed in Equation G.12.

$$\alpha_m = \sigma\sqrt{\frac{n}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} \tag{G.12}$$

(If done properly, a spreadsheet *regression* can give all four of the quantities above.) As in the single data set case, these standard errors should be used

i	x_i	x_i^2	$x_i y_i$	y_i	y_i^2
1	0.1	0.01	0.3	3	9
2	0.2	0.04	0.8	4	16
3	0.3	0.09	1.2	4	16
4	0.4	0.16	2.0	5	25
n	$\sum x_i$	$\sum x_i^2$	$\sum x_i y_i$	$\sum y_i$	$\sum y_i^2$
4	1.0	0.3	4.3	16	66

Table G.1: Sample Data

for the uncertainties in the graphical parameters only for cases with about 30 or more data points. When less data is available, confidence intervals should be used as before. In this case, the 95% confidence interval for the y–intercept is shown in Equation G.13.

$$\Delta b = \sigma\ t_{p,n-2}\sqrt{\frac{\sum x_i^2}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} = t_{p,n-2}\alpha_b \tag{G.13}$$

and the 95% confidence interval for the slope is shown in Equation G.14.

$$\Delta m = \sigma\ t_{p,n-2}\sqrt{\frac{n}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} = t_{p,n-2}\alpha_m \tag{G.14}$$

where p is calculated as before. (The $n-2$ in this case reflects the fact that we have one less degree of freedom since we are estimating two parameters instead of one.)

Following is a calculation of the least squares fit and the standard error of the slope and intercept for some test data. Consider the determination of the 95% confidence intervals to be an exercise for the student.

First

$$n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2 = (4)(0.3) - (1)^2 = 0.2$$

and so

$$b = \frac{\left(\sum y_i\right)\left(\sum x_i^2\right) - \left(\sum x_i\right)\left(\sum x_i y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} = \frac{(16)(0.3) - (1)(4.3)}{0.2} = 2.5$$

thus

$$m = \frac{n\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2} = \frac{(4)(4.3) - (1)(16)}{0.2} = 6.0$$

and

$$S = \sum y_i^2 - m\sum x_i y_i - b\sum y_i = (66) - (6)(4.3) - (2.5)(16) = 0.2$$

$$\sigma = \sqrt{\frac{S}{n-2}} = \sqrt{\frac{0.2}{4-2}} = 0.316228$$

$$\alpha_b = \sigma\sqrt{\frac{\sum x_i^2}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} = (0.316228)\sqrt{\frac{0.3}{0.2}} = 0.3878298$$

$$\alpha_m = \sigma\sqrt{\frac{n}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}} = (0.316228)\sqrt{\frac{4}{0.2}} = 1.414214$$

Thus

$$\Delta b =$$

and

$$\Delta m =$$

and so

$$b = 2.5\pm$$

and

$$m = 6.0\pm$$

If the experimenter has some control over the amount of data taken, the following guidelines should be observed. Determination of the desired number of measurements is constrained by two factors:

1. the number of measurements of each x_i and y_i so that the statistical uncertainties in these are less than the instrumental uncertainties involved,

2. The number of data points n required so that the uncertainties in the parameters will produce an uncertainty in the computed y_i values comparable to the instrumental uncertainties, or

   $$y_i^c = b + mx_i$$

   so

   $$(\Delta y_i^c)^2 = (\Delta b)^2 + x_i^2(\Delta m)^2 + m^2(\Delta x_i)^2$$

   Thus, the computed uncertainty in the computed values y_i^c should be comparable to the actual experimental uncertainties.

Obviously, the determination of n is non-trivial, even in the case of a straight line.

G.4 Propagation of Errors

Rules for combining errors in calculations were given in first year. At that time several were given, but they can all be replaced by the two which follow which are simple once you are familiar with multivariate differentiation. (Actually two slightly different calculations will be shown which will yield different although generally similar results. The difference reflects whether or not different errors are assumed to be independent. Both are shown, and you must learn how to choose which to use in a given situation.) Suppose we have a quantity y which is a function of the m variables $x_1, x_2 \cdots x_m$:

$$y = f(x_1, x_2 \ldots x_m)$$

It can be shown that the **standard error** in y, $\alpha(y)$, is given by:

$$\alpha(y) = \sqrt{\left(\frac{\partial f}{\partial x_1}\Delta x_1\right)^2 + \cdots + \left(\frac{\partial f}{\partial x_m}\Delta x_m\right)^2} = \sqrt{\sum_{i=1}^{m}\left(\frac{\partial f}{\partial x_i}\Delta x_i\right)^2} \quad \text{(G.15)}$$

where Δx_i is the uncertainty in x_i. This equation is valid for uncertainties which are assumed to be *independent*. If errors are independent, then it is unlikely that they will *all* "add up" in their contribution to the total error. The standard error takes this into account (and is sometimes called the **probable error** since an error outside this range is unlikely.) For example, if you measure the diameter and mass of a marble, there is no reason to expect an error in one to affect the other, so when calculating density the standard error should be used. Alternatively, the *maximum* error for a quantity y which is a function of the m variables $x_1, x_2 \ldots x_m$ can be used. It is given by

$$\Delta f(x) = \left|\frac{\partial f}{\partial x_1}\Delta x_1\right| + \cdots + \left|\frac{\partial f}{\partial x_m}\Delta x_m\right| = \sum_{i=1}^{m}\left|\frac{\partial f}{\partial x_i}\Delta x_i\right| \quad \text{(G.16)}$$

This equation is valid for uncertainties which are assumed to be **dependent**. For instance, if you use the same instrument for many measurements, then any systematic error in the instrument will affect all the same way. This could happen if you measured the 3 dimensions of a cube with the same calliper and then calculated the volume. In this case maximum error should be used.

As an example, let us suppose that the volume of a cylinder must be determined from measurements of its length and diameter. The volume is given by:

$$V = \frac{\pi d^2 l}{4}$$

where d is the diameter and l is the length. Suppose 10 measurements of the diameter and length are made, with the following results:

$$d = 5.005 \pm 0.007\text{cm}$$

and

$$l = 10.001 \pm 0.003\text{cm}$$

Then the standard error in the cylinder volume is given by:

$$\begin{aligned} \alpha(V) &= \sqrt{\left(\tfrac{\partial V}{\partial d}\Delta d\right)^2 + \left(\tfrac{\partial V}{\partial l}\Delta l\right)^2} \\ &= \sqrt{\left(\tfrac{\pi dl}{2}\right)^2 \Delta d^2 + \left(\tfrac{\pi d^2}{4}\right)^2 \Delta l^2} \end{aligned}$$

Then the volume of the cylinder is 196.762 cm^3, and the standard error in the volume is 0.554 cm^3, so the result would be stated as

$$V = 196.76 \pm 0.55\text{cm}^3$$

assuming that the 3rd decimal place was uncertain in our original measurements. Calculating the *maximum error* for V using the above data would proceed as follows:

$$\begin{aligned} \Delta V &= \left|\tfrac{\partial V}{\partial d}\right| \Delta d + \left|\tfrac{\partial V}{\partial l}\right| \Delta l \\ &= \left|\tfrac{2\pi dl}{4}\right| \Delta d + \left|\tfrac{\pi d^2}{4}\right| \Delta l \\ &= \left(\tfrac{\pi d^2 l}{4}\right) \tfrac{2\Delta d}{d} + \left(\tfrac{\pi d^2 l}{4}\right) \tfrac{\Delta l}{l} \quad \textit{as in } 1^{st} \textit{ year} \\ &= 0.61\text{cm}^3 \end{aligned}$$

(In this case, the difference between the two uncertainties is small. Had we chosen a different example, the difference may have been more pronounced.)

G.5 References

1. Emerson M. Pugh and George H. Winslow, *The Analysis of Physical Measurements*, Addison-Wesley Series in Physics, 1966

2. J. Topping, *Errors of Observation and Their Treatment*, Chapman and Hall Science Paperbacks, 1972(4th Ed.)

3. Murray R. Speigel, *Statistics*, Schaum's Outline Series in Mathematics, McGraw-Hill, 1961

ν	$t_{.900}$	$t_{.950}$	$t_{.975}$	$t_{.990}$	$t_{.995}$	ν
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
∞	1.282	1.645	1.960	2.326	2.576	∞

Table G.2: Student's T–Distribution

Appendix H

Generalized Least Squares Fitting

H.1 Introduction

Previously you have done curve fitting in two dimensions. Now you will learn how to extend that to multiple dimensions.

H.1.1 Non-linear

Linearizable

Some equations, such as

$$y = Ae^{(Bx)}$$

can be treated fairly simply. Linearize and do a linear least squares fit, as you have done in the past. (Note: "Least Squares" applies to transformed quantities, not original ones so gives a different answer than you would get from a least squares fit in the untransformed quantities; remember in general the idea of a line of "best fit" is not unique. For example, for the equation shown, $\ln y$ vs. x is linear, so you can do a least squares fit in $\ln y$, but this will not give the same result as a least squares fit in y, since the sum of squares of $\ln y$ will depend on the data in a different way than the sum of squares in y.)

H.1.2 Linear

General

Some equations, such as this

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots b_k x_k$$

are linear, although in multiple variables. We can create a matrix of independent data

$$A = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

from the x values, where x_{ij} means variable x_j for data point i and form a vector of dependent data

$$b = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

where y_i is the y data for data point i.

This creates a system which can be solved using the "regression" feature of a spreadsheet. (Be sure to disable the calculation of the y–intercept, as the first coefficient calculated will be the y-intercept, and standard errors will be given for each parameter.)

Polynomial

Consider an equation such as this:

$$y = b_0 + b_1 x + b_2 x^2 + \cdots b_k x^k$$

This is just a special case of the previous situation above, eg. $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$, etc. (or $x_1 = 1/x$, $x_2 = 1/x^2$, etc.)

What about fit with skipped orders?

eg. $y = a + b/x^2 + c/x^5$

In this case, $x_1 = 1/x^2$, $x_2 = 1/x^5$.

H.1.3 Goodness of fit

Often you must choose between different fits because you do not know what type of equation to use. In this case you want to be able to answer the question "Which fit is better?"

1. If both fits have the same number of parameters, then the better fit is the one with the smaller SSE in the *same* quantity. (In other words, if you're comparing a fit in y vs. x to one in $\ln y$ vs. x, you will first have to calculate the SSE of both in y vs. x. If you have linearized an equation to calculate a fit, you can still use that fit to calculate the SSE in the original quantity afterward.)

2. One or both of the fits may have some parameters which are not "statistically significant"; (i.e. lots of parameters close to 0 are probably meaningless.) How close to 0 is "close enough"?

   - RULE: Adding more parameters → smaller SSE, (however a small change in SSE may not be significant.) Whether or not the added parameters are significant can be determined statistically *if the fit is a linear one or one which can be linearized.*

The following example illustrates how to do this for a linear fit.

H.2 Example

Consider the data shown in Table H.1 and plotted in Figure H.1. (Error bars have been omitted for simplicity.)

It should be obvious that some possible equations for a fit to this data may be polynomials in $1/x$.

Some of these are shown in Figures H.2, H.3, and H.4.

1. Do fit with $g + 1$ parameters (as above); calculate the sum of squares error and call it SSE_1. If you use regression from a spreadsheet, you can determine SSE from the results. Remember $SSE_1 = {s_1}^2\nu_1$; in this case $\nu_1 = n - (g + 1)$.

   This gives us Table H.2 and the resulting graph is Figure H.2

   Notice that the curve cannot "bend enough", and so we will see what happens if we add another parameter.

x	y
100	1
85	2
70	4
50	8
36	15
20	25
10	45

Table H.1: Sample Data

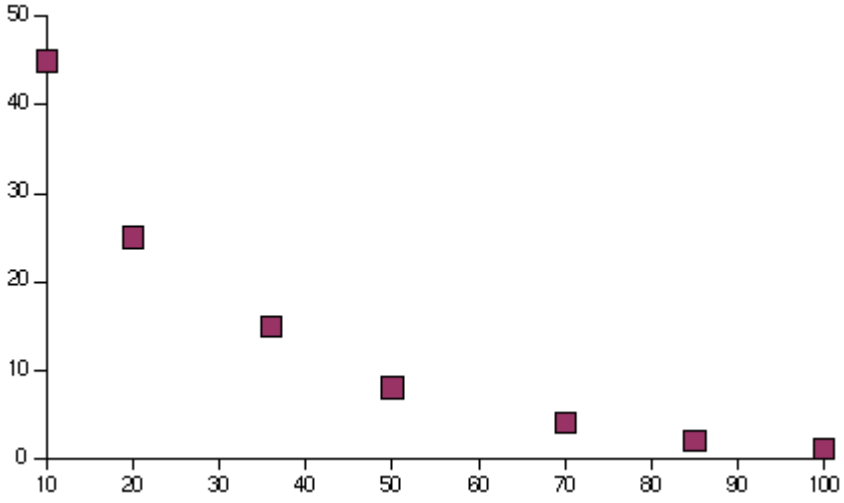


Figure H.1: Plot of Sample Data

$1/x$	y
0.0100	1
0.0118	2
0.0143	4
0.0200	8
0.0278	15
0.0500	25
0.1000	45

Table H.2: Data for Fit to $A + B/x$

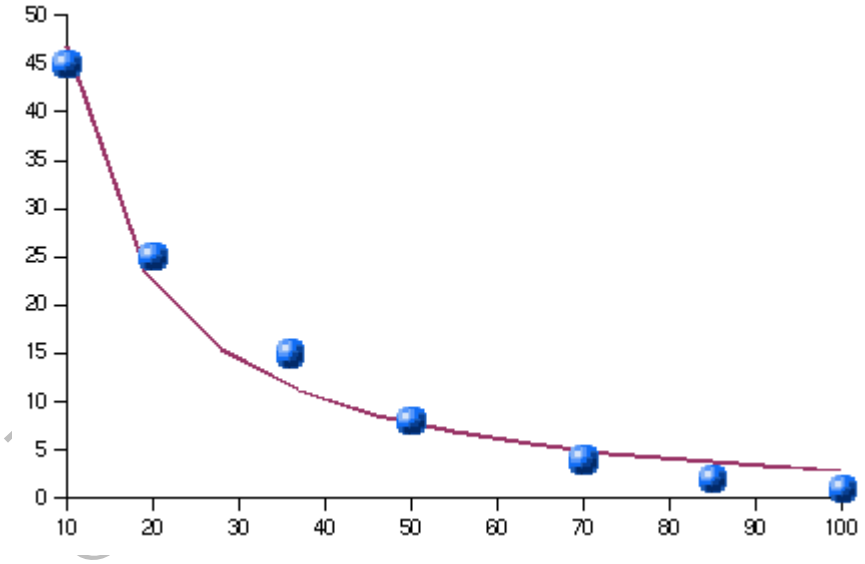


Figure H.2: Graph of Fit to $A + B/x$

$1/x$	$1/x^2$	y
0.01000	0.000100	1
0.01176	0.000138	2
0.01429	0.000204	4
0.02000	0.000400	8
0.02778	0.000772	15
0.05000	0.002500	25
0.10000	0.010000	45

Table H.3: Data for Fit to $A + B/x + C/x^2$

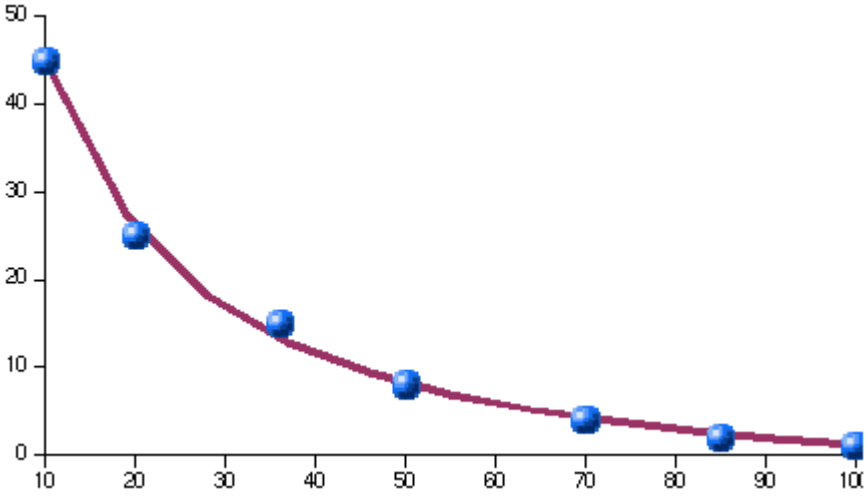


Figure H.3: Graph of Fit to $A + B/x + C/x^2$

In our case, to compare the 2 parameter fit to the three parameter fit we do this by creating a matrix

$$A_1 = \begin{pmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_n \end{pmatrix}$$

and we solve as described earlier.

2. Do fit with $k+1$ parameters (as above); calculate SSE_2. As above, $SSE_2 = {s_2}^2\nu_2$ and in this case $\nu_2 = n-(k+1)$.

   In our case, we do this by creating a matrix

$$A_2 = \begin{pmatrix} 1/x_1 & 1/x_1^2 \\ 1/x_2 & 1/x_2^2 \\ \vdots & \vdots \\ 1/x_n & 1/x_n^2 \end{pmatrix}$$

   and repeat.

3. Calculate s_3 as follows:

$$s_3 = \sqrt{\frac{SSE_1 - SSE_2}{k-g}}$$

   and let $\nu_3 = k-g$.

4. Calculate F as follows:

$$F = \frac{{s_3}^2}{{s_2}^2}$$

   If F is *big*, then include the extra parameters. (In this case, it means the SSE changed a lot by adding the extra parameters, which is what would happen if they were really important.) How big is "big"?

5. Look up F_{α,ν_3,ν_2} from a table of the F distribution in a statistics text, where α determines the confidence interval; typically $\alpha = 0.05$ for a 95% confidence interval. If the F you calculated is *greater* than the table value, then keep the extra parameters. *Note: In the table, you*

$1/x$	$1/x^2$	$1/x^3$	y
0.01000	0.000100	0.00000100	1
0.01176	0.000138	0.00000163	2
0.01429	0.000204	0.00000292	4
0.02000	0.000400	0.00000800	8
0.02778	0.000772	0.00002143	15
0.05000	0.002500	0.00012500	25
0.10000	0.010000	0.00100000	45

Table H.4: Data for Fit to $A + B/x + C/x^2 + D/x^3$

are given quantities ν_1 and ν_2; you should use your calculated value of ν_3 in place of ν_1 in the table. Doing it this way keeps the table in the same form you will find it in a statistics text.

(Note that in some of the figures, the fit curve is not quite smooth, due to an insufficient number of plotting points used.) It is not immediately obvious which of the above curves fits the data "best". We could even go on adding higher and higher powers of $1/x$ until we had no more *degrees of freedom* [1] left, but once we get no significant change, it's time to stop.

Usually we want to compare two fits; in this example, we will compare 3 fits to illustrate the process more clearly. We will compare 2 fits at a time, and in each case we will use $g+1$ [2] to denote the number of parameters in the "smaller" fit, and $k+1$ to denote the number of parameters in the "bigger" fit, so k is always bigger than g.

[1] The number of degrees of freedom in a fit is the number of data points beyond the bare minimum for that fit. So, for an average it is $n-1$, since only one value is needed; for a straight line it is $n-2$, since two points are needed, etc. In general,

$$\nu = n - m$$

where m is the number of parameters in the fit to be determined. Note that when you have no degrees of freedom, you have no idea of the "goodness" of your data, and thus cannot determine the standard deviation. Once you have even one degree of freedom, you can do so.

[2] Why not just g? Because g is the *degree* of the polynomial, which has $g+1$ parameters. For example a polynomial of degree 2, such as $Ax^2 + Bx + C$ has 3 parameters, namely A, B, and C.

Quantity	Value $A + B/x$	Value $A + B/x + C/x^2$	Value $A + B/x + C/x^2 + D/x^3$
s	2.43	1.04	0.78
ν	5	4	3
SSE	29.43	4.31	1.83
A	-2.02	-6.45	-9.93
σ_A	1.37	1.09	1.91
B	488.15	784.5	1156.6
σ_B	30.56	62.79	189.98
C		-2713.14	-12121.37
σ_C		562.24	4672.18
D			60476.14
σ_D			29909.81

Table H.5: Comparisson of Fit Parameters

- Note that the SSE gets smaller as the number of parameters increases, but the change gets smaller.
- Note also that when a parameter is added, all of the previous parameters change as well.
- Even though it is considered "insignificant", the D parameter is bigger than all of the rest! (However, note the size of its standard error. Remember also that it gets divided by x^3, which will range in size from $1000 \rightarrow 1000000$.)

ν_2 \ ν_1	1	2	3	4	5	6
1	161.4	199.5	215.7	224.6	230.2	234
2	18.51	19	19.16	19.25	19.3	19.33
3	10.13	9.55	9.28	9.12	9.01	8.94
4	7.71	6.94	6.59	6.39	6.26	6.16
5	6.61	5.79	5.41	5.19	5.05	4.95
6	5.99	5.14	4.76	4.53	4.39	4.28
7	5.59	4.74	4.35	4.12	3.97	3.87
8	5.32	4.46	4.07	3.84	3.69	3.58
9	5.12	4.26	3.86	3.63	3.48	3.37
10	4.96	4.1	3.71	3.48	3.33	3.22
11	4.84	3.98	3.59	3.36	3.2	3.09
12	4.75	3.89	3.49	3.26	3.11	3
13	4.67	3.81	3.41	3.18	3.03	2.92
14	4.6	3.74	3.34	3.11	2.96	2.85
15	4.54	3.68	3.29	3.06	2.9	2.79
16	4.49	3.63	3.24	3.01	2.85	2.74
17	4.45	3.59	3.2	2.96	2.81	2.7
18	4.41	3.55	3.16	2.93	2.77	2.66
19	4.38	3.52	3.13	2.9	2.74	2.63
20	4.35	3.49	3.1	2.87	2.71	2.6
21	4.32	3.47	3.07	2.84	2.68	2.57
22	4.3	3.44	3.05	2.82	2.66	2.55
23	4.28	3.42	3.03	2.8	2.64	2.53
24	4.26	3.4	3.01	2.78	2.62	2.51
25	4.24	3.39	2.99	2.76	2.6	2.49
26	4.23	3.37	2.98	2.74	2.59	2.47
27	4.21	3.35	2.96	2.73	2.57	2.46
28	4.2	3.34	2.95	2.71	2.56	2.45
29	4.18	3.33	2.93	2.7	2.55	2.43
30	4.17	3.32	2.92	2.69	2.53	2.42
40	4.08	3.23	2.84	2.61	2.45	2.34
60	4	3.15	2.76	2.53	2.37	2.25
120	3.92	3.07	2.68	2.45	2.29	2.17
∞	3.84	3	2.6	2.37	2.21	2.1

Table H.6: F–Distribution Table ($\alpha = 0.05$)

ν_2 \ ν_1	7	8	9	10	12	15	20	24
1	236.8	238.9	240.5	241.9	243.9	245.9	248	249.1
2	19.35	19.37	19.38	19.4	19.41	19.43	19.45	19.45
3	8.89	8.85	8.81	8.79	8.74	8.7	8.66	8.64
4	6.09	6.04	6	5.96	5.91	5.86	5.8	5.77
5	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53
6	4.21	4.15	4.1	4.06	4	3.94	3.87	3.84
7	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41
8	3.5	3.44	3.39	3.35	3.28	3.22	3.15	3.12
9	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.9
10	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74
11	3.01	2.95	2.5	2.85	2.79	2.72	2.65	2.61
12	2.91	2.85	2.8	2.75	2.69	2.62	2.54	2.51
13	2.83	2.77	2.71	2.67	2.6	2.53	2.46	2.42
14	2.76	2.7	2.65	2.6	2.53	2.46	2.39	2.35
15	2.71	2.64	2.59	2.54	2.48	2.4	2.33	2.29
16	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24
17	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19
18	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15
19	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11
20	2.51	2.45	2.39	2.35	2.28	2.2	2.12	2.08
21	2.49	2.42	2.37	2.32	2.25	2.18	2.1	2.05
22	2.46	2.4	2.34	2.3	2.23	2.15	2.07	2.03
23	2.44	2.37	2.32	2.27	2.2	2.13	2.05	2.01
24	2.42	2.36	2.3	2.25	2.18	2.11	2.03	1.98
25	2.4	2.34	2.28	2.24	2.16	2.09	2.01	1.96
26	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95
27	2.37	2.31	2.25	2.2	2.13	2.06	1.97	1.93
28	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91
29	2.35	2.28	2.22	2.18	2.1	2.03	1.94	1.9
30	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89
40	2.25	2.18	2.12	2.08	2	1.92	1.84	1.79
60	2.17	2.1	2.04	1.99	1.92	1.84	1.75	1.7
120	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61
∞	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52

Table H.7: F–Distribution Table ($\alpha = 0.05$)

ν_2 \ ν_1	30	40	60	120	∞
1	250.1	251.1	252.2	253.3	254.3
2	19.46	19.47	19.48	19.49	19.5
3	8.62	8.59	8.57	8.55	8.53
4	5.75	5.72	5.69	5.66	5.63
5	4.5	4.46	4.43	4.4	4.36
6	3.81	3.77	3.74	3.7	3.67
7	3.38	3.34	3.3	3.27	3.23
8	3.08	3.04	3.01	2.97	2.93
9	2.86	2.83	2.79	2.75	2.71
10	2.7	2.66	2.62	2.58	2.54
11	2.57	2.53	2.49	2.45	2.4
12	2.47	2.43	2.38	2.34	2.3
13	2.38	2.34	2.3	2.25	2.21
14	2.31	2.27	2.22	2.18	2.13
15	2.25	2.2	2.16	2.11	2.07
16	2.19	2.15	2.11	2.06	2.01
17	2.15	2.1	2.06	2.01	1.96
18	2.11	2.06	2.02	1.97	1.92
19	2.07	2.03	1.98	1.93	1.88
20	2.04	1.99	1.95	1.9	1.84
21	2.01	1.96	1.92	1.87	1.81
22	1.98	1.94	1.89	1.84	1.78
23	1.96	1.91	1.86	1.81	1.76
24	1.94	1.89	1.84	1.79	1.73
25	1.92	1.87	1.82	1.77	1.71
26	1.9	1.85	1.8	1.75	1.69
27	1.88	1.84	1.79	1.73	1.67
28	1.87	1.82	1.77	1.71	1.65
29	1.85	1.81	1.75	1.7	1.64
30	1.84	1.79	1.74	1.68	1.62
40	1.74	1.69	1.64	1.58	1.51
60	1.65	1.59	1.53	1.47	1.39
120	1.55	1.5	1.43	1.35	1.25
∞	1.46	1.39	1.32	1.22	1

Table H.8: F–Distribution Table ($\alpha = 0.05$)

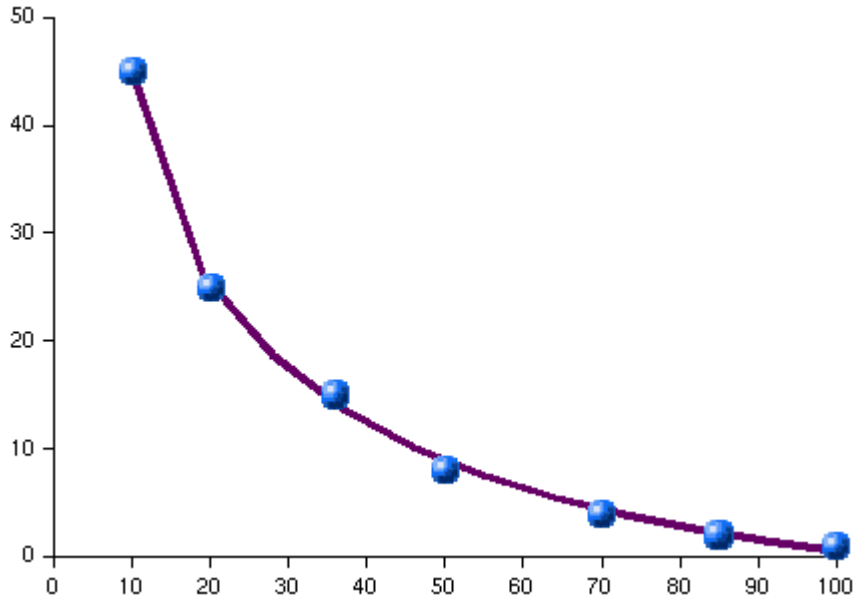


Figure H.4: Graph of Fit to $A + B/x + C/x^2 + D/x^3$

Appendix I

Determining Graph Scatter

I.1 Introduction

You may have wondered if it's possible to determine automatically whether a linear graph has a "small" or a "large" scatter. Remember that the scatter is considered "small" if a line can be drawn that crosses the error bars for each data point. If no such line can be drawn, the scatter is considered "large". Consider a data point near the line of best fit. A point is considered "close"

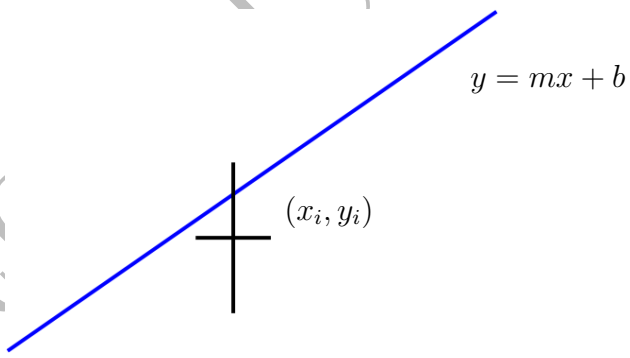


Figure I.1: Data point near best fit line

to the line if it passes within the box formed by the error bars for the point.

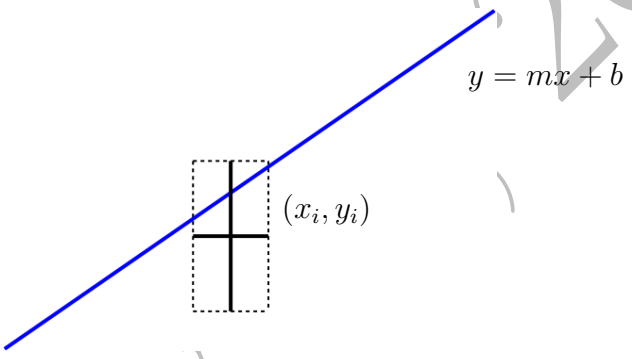


Figure I.2: Data point near best fit line with error "box"

Note that it's possible that a line could cut through one corner of the box without actually crossing one of the error bars.

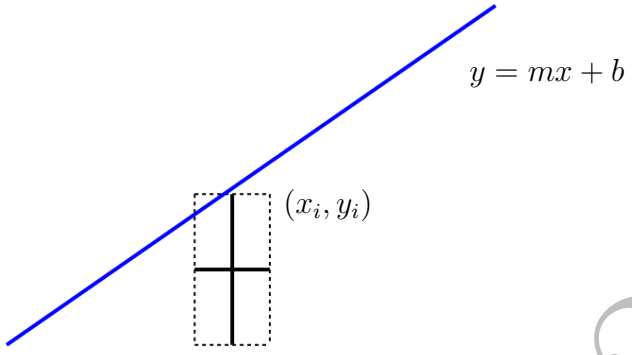


Figure I.3: Line inside "box" but missing error bars

Let's take a closer look at a data point. For point i, the error bars form a box with two diagonals.

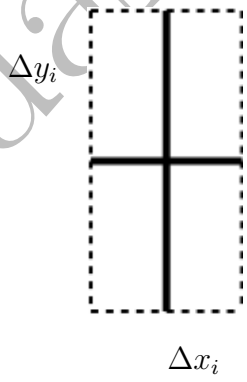


Figure I.4: Data point close-up view

We can redraw the "error box" showing the diagonals.

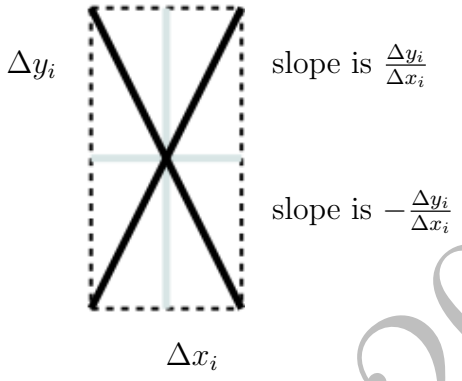


Figure I.5: Diagonals from error bars

Now it should be clear that if a line passes through this box, *even if it doesn't cross one of the error bars*, then it must cross one of these diagonals with an x value between $x_i + \Delta x_i$ and $x_i - \Delta x_i$.

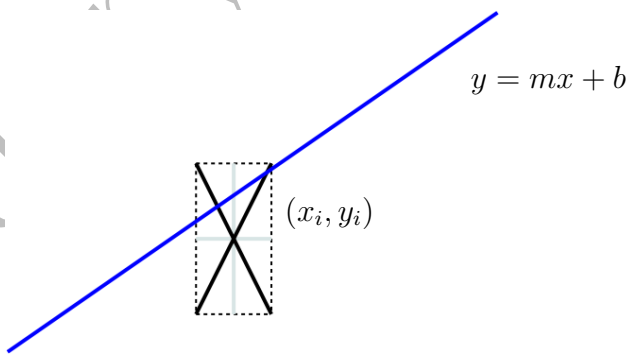


Figure I.6: Test whether line passes within error bars

The slope of the diagonal with the positive slope is

$$m_{+i} = \frac{\Delta y_i}{\Delta x_i} \tag{I.1}$$

The y-intercept of the diagonal with the positive slope is

$$b_{+i} = y_i - m_{+i} x_i \tag{I.2}$$

since, by definition, (x_i, y_i) must be on the line. Similarly, The slope of the diagonal with the negative slope is

$$m_{-i} = -\frac{\Delta y_i}{\Delta x_i} \tag{I.3}$$

The y-intercept of the diagonal with the negative slope is

$$b_{-i} = y_i - m_{-i} x_i \tag{I.4}$$

since, by definition, (x_i, y_i) must be on the line. For two lines $y = m_1 x + b_1$ and $y = m_2 x + b_2$, they cross when there is a single point (x_c, y_c) that satisfies both equations. Thus

$$y_c = m_1 x_c + b_1$$

and

$$y_c = m_2 x_c + b_2$$

Combining this gives

$$m_1 x_c + b_1 = m_2 x_c + b_2$$

and so

$$x_c = \frac{b_2 - b_1}{m_1 - m_2} \tag{I.5}$$

To see where the best fit line $y = mx + b$ crosses one of the diagonal lines, we just use Equation I.5 and so

$$x_{ci+} = \frac{b - b_{+i}}{m_{+i} - m} \tag{I.6}$$

and

$$x_{ci-} = \frac{b - b_{-i}}{m_{-i} - m} \tag{I.7}$$

If either of x_{ci+} or x_{ci-} is between $x_i + \Delta x_i$ and $x_i - \Delta x_i$ then the best fit line crosses the error bars for point i. If this is true for *all* of the data points, then this is a case of small scatter.

I.1.1 When this doesn't work

This tests whether a *given* line crosses all of the error bars. So, if the least squares fit line crosses all of the error bars, then this test is easy to use. It's possible, though, that the least squares fit line *doesn't* cross all of the error bars, but another line can be drawn which does. Unless we can determine the equation of that line, we can't apply this test.

Another case where this doesn't work

If the x co-ordinate has no uncertainty, (i.e. there are only error bars in the y direction), then this won't work either. However, in that case, it easy to apply a similar test to see if the line crosses all of the (vertical) error bars. In this case, for each x_i, it must be that

$$y_i - \Delta y_i \leq mx_i + b \leq y_i + \Delta y_i$$

or

$$|y_i - (mx_i + b)| \leq \Delta y_i$$

Appendix J

Lab Checklist

This marking checklist will be used for lab reports this term. You need to print one off and attach it to each lab report you hand in. Lab reports will be marked as follows:

- Start with 90

  For items *not* in italics

- Subtract 1 for each ˜.
- Subtract 2 for each − .

  For items *in* italics

- Subtract 3 for each ˜.
- Subtract 6 for each − .

Note the importance of items in italics. These are very important in a report, and so are weighted accordingly.
The other 10 marks will be based on how well the post-lab discussion questions were answered within the text of the report. *Remember that the answers to these questions should be an integral part of the report, not merely an afterthought.*

Lab Format Checklist (V2.25bg)

A. General

1. Your own work _____
2. Complete _____
3. Clear and appropriate "Purpose" _____
4. Flows _____
5. Did not require help on or after due date _____
6. Correct grammar _____
7. Correct spelling _____
8. Complete sentences where required _____
9. Legible _____
10. Professionally presented _____
11. Properly identified (eg. name, partner) _____
12. On time _____
13. Checklist included _____
14. Template included _____

B. Data (for data not in tables)

1. Your own data _____
2. Values recorded with uncertainties _____
3. Sufficient data _____
4. Reasonable values _____
5. Reasonable uncertainties _____
6. Correct number of significant figures _____
7. Units recorded _____

C. Data in Tables

1. Neat _____
2. Column headings informative _____
3. Units given _____
4. Uncertainties given _____
5. Label _____
6. Number given (eg. "Table #2") _____

D. Calculations and Results

1. Any required derivations done correctly _____
2. Analysis explained where needed _____
3. Correct formulas used _____
4. Sample calculations shown where needed _____
5. All required values calculated _____
6. Uncertainties included _____
7. Units included _____
8. Correct number of significant figures _____
9. Appropriate use of standard form _____
10. Theoretical or reasonable value _____
11. Agreement of experiment with theory _____

E. Graphs

1.	Title meaningful	_____
2.	Correct graph type and orientation	_____
3.	Background colour and grid appropriate	_____
4.	Plotting data in table	_____
5.	Axis labels meaningful	_____
6.	Correct axis units	_____
7.	Points not connected	_____
8.	Error bars in both dimensions or note if too small	_____
9.	Error bars correct size	_____
10.	Line of best fit shown without markers	_____
11.	Number given (eg. "Graph #3")	_____

F. Least Squares Fits

1.	Points used for fit clearly identified	_____
2.	Results given meaningful names	_____
3.	Correct units for slope and intercept	_____
4.	Correct indication of "large" or "small" scatter	_____

G. *Error Discussion*

1.	*Sources listed are significant*	_____
2.	*Sources are prioritized*	_____
3.	*Systematic error consequences*	_____
4.	*Evidence: ie test or bound*	_____
5.	*Reasonable suggestions for improvement*	_____

H. *Conclusions*

1. *Relate to purpose* _____
2. *Major results stated* _____
3. *Comparisons made where appropriate* _____
4. *Agreement noted when found* _____
5. *% difference calculated only when no agreement* _____

I. References

1. Source(s) of constants listed _____

J. Methods

1. All steps clearly described _____
2. Paragraph format _____
3. Past tense _____

K. *Introduction*

1. *Rationale for research given* _____
2. *Historical/experimental context given* _____
3. *Good references* _____

L. Abstract

1. Important results summarized _____
2. Paragraph format _____
3. Appropriate verb tenses _____

Index