Lesson 1: Introduction to Uncertainty

I am confident that everyone taking this course has done a scientific word problem before. More than likely you have done many such problems in your life. With these questions you are given values for several different physical parameters and are asked to find the value of some unknown quantity. In doing all of these questions you have become very skilled in finding *the answer*. There is no debate allowed, no reason to doubt the number that comes up on your calculator. You followed the proper steps, found the solution to the question asked, and everyone in the world should be able to come up with the exact same answer.

However this kind of infinite precision in your answers becomes impossible to have when we start doing physics in the real world. It is the goal of this manual to introduce you to the ways in which modern science deals with this issue. We will discuss ways of estimating uncertainty in measurements, translating those into the uncertainty in results, and how to properly present results with that uncertainty. In your assignments you will practice the basics over and over again so that you become confident in their usage. In addition, you will be exposed to some more advanced techniques. Mastery of these sophisticated topics is not expected, but becoming familiar with them now will enhance your experimental experiences in the future.

Uncertainty in Measurement

In textbook problems you are able to have complete certainty in your answer because the values you start with are exact as well. If a mass of 5 kg is accelerating at a rate of 3 m/s^2, then the net force on the mass must be *exactly* 15 Newtons. If we try and conduct an experiment to test this we will have to measure the mass and the acceleration and those measurements will never be perfect. We will always have some sort of uncertainty in our measurements, and therefore we will have uncertainty in our results.

One rule that many people learn is that the uncertainty in the measurement is half of the smallest scale marking on the measuring device. For example the uncertainty of a meter stick that is marked every millimeter would be 0.5mm. However, we cannot rely on such a simple rule. We have to take more things into account about how we measure. Would we expect a person with good vision to have the same uncertainty in their ruler measurement as someone who could not see as well? Would we expect that the uncertainty in measuring the distance between two thin pencil lines is the same as measuring the distance between two lines made with a broad marker?

Many times the best practice is to make the measurement several times and create an error profile for that measurement. Later in this manual we will look at the process of transferring the uncertainty in what is directly measured to the an uncertainty for a result. This can be a tedious process. Most of the time it is best to simply conduct the entire experimental procedure a number of times and then run statistics on your set of answers. It is with this in mind that the statistical procedures are developed from here on.

Because we will be taking this approach we will be very liberal in our use of the word measurement; it will not always refer to the direct measurement of a variable. If an experiment is said to measure a value of g then it is likely that other variables are directly measured and g is the final output of the experiment. However it is common usage to refer to the experiment as measuring g.

Significant Figures & Finite Precision

Any experimental result should be presented with some form of uncertainty. The most basic form of uncertainty analysis is the use of significant figures. In the last section we talked about how measurements can only be known to a certain decimal place. The significant figure system of uncertainty gives us rules that will tell us the number of decimal places we know the result to given the numbers we have for the measurements. Before we give the rules for showing how this system works, we must first define what a significant figure is.

- All non-zero digits are significant.
- Leading zeros are never significant.
- Imbedded zeros (for example, 506 kg) are always significant.
- Trailing zeros (those to the right of non-zero digits are significant only if the decimal point is shown.
- Counting numbers (for example, 11 apples or 5 bricks) are known exactly; they have an infinite number of significant figures.

Here are some examples:

– 0.00334 m ---------- 3 sig. figs

– 1.0058 kg ---------- 5

– 500 grams ---------- 1

– 500. grams ---------- 3

– 34.00 volts ---------- 4

– 0.678 Newtons ---------- 3

– 1.38×10^4 seconds ---------- 3

– 1.3800×10^4 seconds -------- 5

– 1 meter ---------- 1

– 1 ruler ---------- infinite

If the *only* calculations preformed are the addition or subtraction of numbers, the last significant figure in the result occurs at the last decimal place that all numbers in the sum had a significant digit.

$$\begin{array}{r} 8.16\ \text{m.} \\ +\ 74\quad\ \text{m.} \\ \hline 82\quad\ \text{m.} \end{array}$$

When multiplying or dividing numbers, the number of significant digits in the result equals the smallest number of significant figures in *any* of the original numbers.

$$8.3\ \text{m} \times 1{,}045\ \text{N} = 8{,}700\ \text{N} \cdot \text{m}$$

There are almost no calculations which involve only addition or subtraction, so the rule for multiplying or dividing will be used in all but very rare cases. Perform your calculations carrying all of the digits, then round the final result to the correct number of significant figures.

This last step may be more tricky that it seems. In an experimental setting you are constantly making calculations and it is possible that you will need to report

any of them. This is part of the reason that computers have become such a large part of modern science. Spreadsheet programs such as Excel allow us to store all of our work and to make sure all calculations are done from the original data. When determining the number of sig. figs. to use you cannot look at the numbers that went directly into the calculation. You must look at all of the raw data that would have an effect on that calculation, even if the raw data was used many steps ago.

Simple Error Calculations

For basic calculations where no statistical analysis is performed we can use some simple formulas to evaluate results: Percent Error, Percent Difference, and Percent Change.

Percent Error is used when comparing an experimental result to an agreed upon standard or a theoretical value.

$$\%\ \text{error} = \left| \frac{\text{standard value - experimental value}}{\text{standard value}} \right| \times 100$$

Percent Difference is used to compare two experimental results of the same quantity that were obtained through different methods.

$$\%\ \text{difference} = \left| \frac{\text{difference of experimental values}}{\text{average of experimental values}} \right| \times 100$$

Percent Change allows the experimenter to compare before and after results of a variable. Please note that the absolute value is not used here so that the calculation can infer if the change was positive or negative.

$$\%\ \text{change} = \frac{\text{final value - initial value}}{\text{initial value}} \times 100$$

Formatting Uncertainty Results

When moving beyond the simple sig. fig. method, the best practice is to quote a plus-or-minus value along side your result. For example, claiming an event lasted 2.53 ± .03 seconds means that you're reasonably confident the event took between 2.50 seconds and 2.56 seconds.

In general, a plus-or-minus value should be quoted to <u>one</u> significant digit. Claiming an uncertainty of ± .036 seconds is absurd…who cares about uncertainty in the thousandths place when we're not even sure what's going on in the hundredths place? That uncertainty is properly reported as ± .04 seconds.

Then the actual measurement should be rounded off to the same place value as the plus-or minus value…since that's where the error starts to creep in. Thus:

2.835 ± .08 seconds is wrong
2.835 ± .082 seconds is wrong
2.84 ± .08 seconds is right

Note that the uncertainty must be fully calculated and formatted before you can determine how to record the answer. This is a little unusual as you must determine how to format what is on the right of the plus-minus sign before you determine how to record the what is on the left.

One other thing to note is the treatment of order of magnitude and units. The cleanest way to present this is to treat any power of ten as part of the unit and place both at the end of the uncertainty statement as in the example below.

$1.78 \pm 0.07 \times 10^{13}$ kg

Types of Error

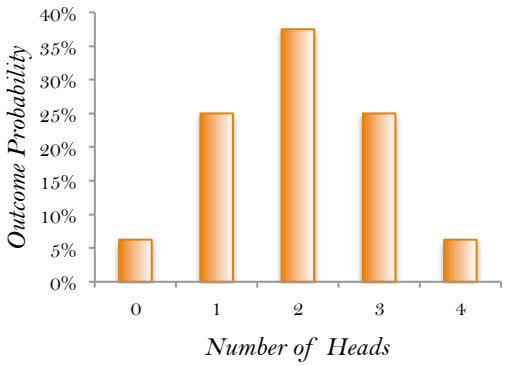
Before we discuss how we are going to quantify uncertainty, we must make sure we understand exactly what we are attempting to quantify. Errors in experimentation fall into two broad categories: systematic error and random error. Systematic errors are those that tend to bias the results in one particular direction. The main causes tend to be a lack of proper planning and design as well as failure to calibrate devices at the beginning of an experiment. One example might be if a scale was not calibrated to read zero when no weight was on it. Then all mass readings would be either all high or all low depending on which way the scale was unbalanced. On the other hand, random errors arise from things that are much harder to control for. A common source of random error is human error. We can make simple mistakes such as misreading a measuring device or disturbing the experiment, but the key thing to remember here is that those types of errors will be random in nature. It is just as likely that we will make a mistake that gives a higher value rather than a lower value.

All of the calculation techniques that we look at this semester will be an attempt to quantify random error. The lack of focus on systematic error is not because it is unimportant, but rather that these types of error are best dealt with before (and during) data collection. If we take care in our experimental design and make sure to calibrate our instruments we can eliminate systematic error before we take our data. Once we have our results we can assume that any uncertainty remaining is due to random error. This will be the approach we will take in this course, but make sure you understand this is an assumption. There are many techniques that are used to analyze the error analysis itself in order to determine if systematic effects were not taken into account. We leave it to others to pass this knowledge along to you.

Random Error Distribution

In statistics, a quantity that is driven by a random process is known as a random variable. A simple example would be an experiment where you flip four coins simultaneously and count how many coins are showing 'heads.' In this case the random variable would be the number of coins showing heads and the random process would be the idea that for each coin there is a 50/50 chance of showing heads or tails. There are five possible outcomes to each trial: no coins will show heads, only 1 will, 2 will, 3 will, or all four will. If this experiment is performed a large number of times the results of the experiment can be summarized in the following chart.

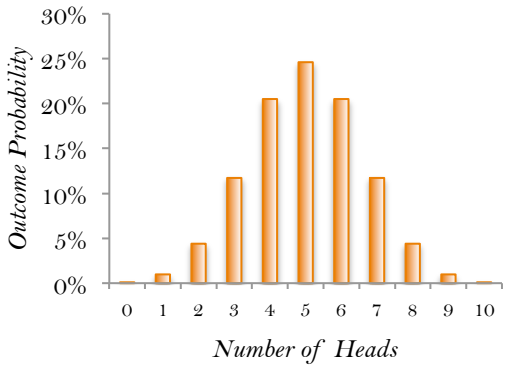
Figure 1
Four Coin Flip Experiment



This distribution should not surprise us. There is a relatively small chance that

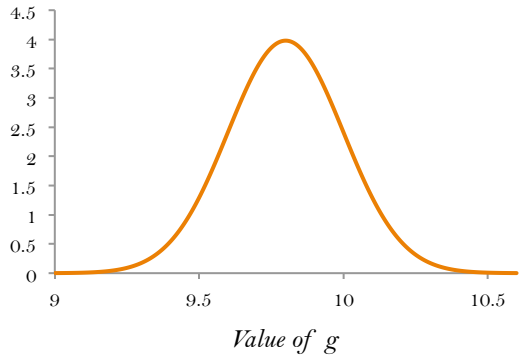
we would get either all heads or all tails, and there is a high probability that we get exactly half the coins to show heads. If we change the experiment to throw 10 coins at once instead of 4 then we get a similar pattern. The chances of getting all heads or all tails is virtually non-existent and the highest probability is still that half of the coins show heads.

Figue 2
Ten Coin Flip Experiment



In this experiment the number of coins showing heads is specifically called a *discrete* random variable. That is only specific values for the variable are possible. There is no way, for example, to have 3.5 coins show heads. A more realistic physics example might be an experiment that attempts to measure the value of *g*. This would be known as a *continuous* random variable as the experiment could produce any value for *g*. In this case the underlying random process would be the fact that there is random error present in all of the measurements that are taken to produce the final result. The distribution for a continuous random variable will simply be an extension of the two charts already shown. Imagine what the graph would look like if 100,000 coins had been tossed at once.

Figure 3
Experiment to Determine *g*



The curve in Figure 3 gives us the probability distribution of a continuous random variable. It is sometimes referred to as *the bell curve* given its shape. However several different distribution curves have a bell shape to them, so this term is ambiguous. The proper term for the function is the Gaussian Distribution. It continues to have the same important properties as the plots for the discrete random variables: there is a high probability for getting a result close to some central value and a low probability of getting a result very far beneath or above the central value. You may notice that the *y-axis* is no longer in percentages and that values for the curve go above 1. This means that we cannot simply find the probability of a particular outcome by knowing the value of the function at some given *x-value*. We will discuss this and the other mathematical values that define this function in the following sections.

Averages

The most obvious feature of the Gaussian Distribution is the peak of the curve around the central value. This

central value is of course the average of all possible values of the variable in question. It should be of no surprise that the average plays a large role in defining a curve that is the consequence of random error. The whole reason that we take an average is to eliminate random error. We know that if we only measure something one time the chances of us getting the exact right value are very low. We feel much more confident in taking the measurement several times and reporting the average value. This increase in confidence is because we have eliminated some of the random error from our result.

As you may know there are several different ways to calculate the average value of a data set. The most common are *the median, the mode, and the mean*[1]. Consider the following data set.

$$\{1, 2, 2, 3, 4, 7, 9\}$$

The median is defined as the middle number of the set when they are arranged in numerical order as above. The median of this set is 3. If a set has an even number of data in it then the median will be the value that is in-between the two middle numbers.

The mode of a set is the number that appears most often in a set. In the example above the mode is 2.

The mean is the most common way to calculate the average. It is the sum of all numbers in the set divided by the number of values in the set. For our sample set, the mean is 4. The formal way of writing the mean is,

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Please note that we will use the notation of an x with a bar over it for an average value. This is sometimes pronounced 'x bar'. It is also important to note that the mean, median, and mode of a Gaussian curve will be the same value. This is not true for all probability curves.

Spread in the Data

Of course the average is not the only number that can describe a data set. It would be possible to have many different data sets with the same average but that are very different from each other. Take the following two sets as examples.

$$\textit{Set A} \quad \{98, 100, 102\}$$
$$\textit{Set B} \quad \{0, 100, 200\}$$

Both of these sets have an average value of 100, but if these were results from an experiment, we would certainly feel that the data from *Set A* was much more reliable than the data from *Set B* due to the differences in spread of the two sets. The most common way of calculating such a spread in the data is the *standard deviation*, which is defined as[2],

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

For our two sets we would have,

[1] There are actually several different types of means as well, the arithmetic mean and the geometric mean being the most common. Throughout this text when we say mean we are referring to the arithmetic mean.

[2] Some of you may have seen a formula where the denominator is $N - 1$. We will discuss this difference later on.

$$S_{Set\ A} = \sqrt{\frac{(-2)^2 + (0)^2 + (2)^2}{3}} = \sqrt{\frac{8}{3}} = 1.633$$

$$S_{Set\ B} = \sqrt{\frac{(-100)^2 + (0)^2 + (100)^2}{3}} = \sqrt{\frac{20{,}000}{3}} = 81.65$$

The idea is to take the difference between each data point and the average of the set, square all of the differences, add the squared values together, divide by N, and then take the square root. The differences are squared so that negative deviations will not cancel out positive ones[3]. Taking the square root will ensure that the standard deviation will have the same units as the original data, which makes them easier to compare. If you do not take the square root, then you have a quantity called the *variance.*

$$variance = S^2$$

Our two examples are small data sets, but there is an important feature to the standard deviation that occurs with larger data sets. If we have an experiment that we perform many times, the standard deviation will be the same no matter how many times the experiment is repeated. That is the formula above is not dependent on N. At first this may seem counter intuitive, but let's take a closer look. As more trials are performed N will go up, so the denominator will get larger. However, each time another trial is conducted the sum in the numerator will get an additional term. Once more this additional term will be positive, (or 0), as it is always squared. Most of the time the numerator will grow only a little as most trials will be near the average. However, occasionally, a large deviation will occur that will offset all of the small deviations. In this way both the numerator and denominator grow linearly with N, and so N will have no overall effect.

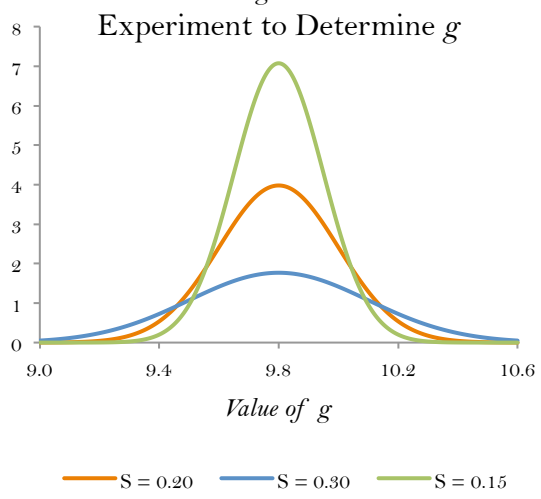
Remember that the goal of doing these calculations is to quantify random error that is present in your experiment. The fact that standard deviation is not dependent on the number of trials you do fits well with this idea. If you change nothing about your experiment as you repeat trials then you face the same random error in your final trial as you did in your first trial. This is a great qualitative way to think of standard deviation. It is an estimation of random error present in any one trial of the experiment.

Standard Deviation & the Gaussian Distribution

The standard deviation will also help define the Gaussian Distribution and the relationship between the two will help us make sense of our uncertainty results. Figure 4 below shows three different Gaussian curves all with a peak value of 9.8, but with different standard deviations.

[3] You may ask why not just take the absolute value? In the age of modern computing this may be a more viable option than it was in the past. Remember that the absolute value function is not differentiable at $x = 0$ whereas the squared function is. This makes more complex calculations easier with the above formula. Also there are times when the variance is a useful number to have. Some will argue that the absolute value method is better, you can research this and decide on your own.

Figure 4



The change in standard deviation causes a change in the width of the curve. The smaller the deviation, the more narrow the peak. In order to interpret these different curves, it is important to understand how we get probabilities off of the curves. Earlier we said that the *y-axis* values were not percentages. Part of the reason for this is that each of the curves in Figures 3 and 4 are normalized so that the area underneath the curve is exactly equal to 1. The way to interpret this is to say that the probability that a single trial of the experiment will fall between ∞ and $-\infty$ is 100%. This is a true, but not very illuminating statement. However, similar statements can be made with different boundaries. If I wanted to know the probability of finding a result between 9.4 and 10.2 in one of the curves in Figure 4, then I would need to find the area under the curve between those two points. This can of course be done with integration, but luckily the results are quite standard and are summarized below.

Table 1

% of trials that lie within…	… standard deviations of the average value
68.27	1.00
95.00	1.96
99.00	2.58
99.73	3.00
99.994	4.00
99.99994	5.00

For example, if you are performing the experiment characterized by the orange line in Figure 4, then 95% of the trials you do will lie within ± 1.96 standard deviations of the average. That is 95% of your trials will lie within ± 0.392, (1.96 * 0.20 = 0.392), of 9.8 or 95% of your trials will lie between 9.408 and 10.192.

This same statement can be reversed. If another trial is performed then you can be 95% sure that it will fall within that same range. This is an interesting statement, but it doesn't quite get us to what we want. If you performed this experiment, you would want to report your result as g = 9.8, but you need to also publish what the uncertainty in that average value is. The standard deviation cannot do this. It can only tell you about the random error associated with a single trial of the experiment.

Standard Error

Let's review a few thoughts before we continue. We know that it is better to perform multiple trials and report the average of those trials rather than only doing it once. We know that this is because random error is reduced in the process of averaging. This should mean that any estimation of uncertainty in the average should be less than the uncertainty in any given trial. We would also feel more confident about an

average that came from a large data set compared to a small one. The calculation of *standard error* will fit these ideas quite well.

$$\overline{S} = \frac{S}{\sqrt{N}}$$

The standard error is simply the standard deviation divided by the square root of the number of trials performed. This formula will ensure that the standard error is always less than the standard deviation and that as the number of trials goes up, the standard error will be reduced. Therefore the correct way to calculate uncertainty for an average is to use the standard error. For this reason the standard error is also referred to as the *standard deviation of the mean*.

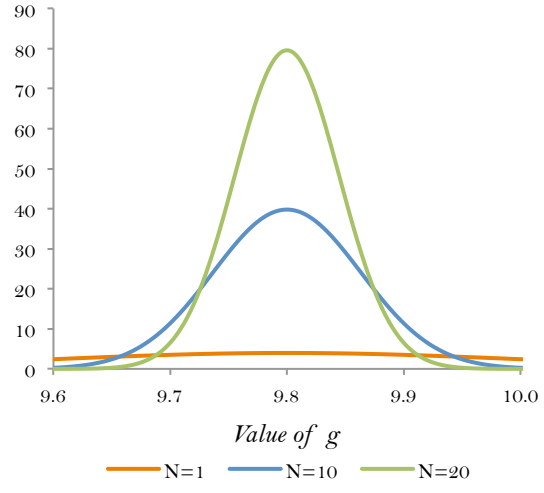
Now these properties could be obtained if the denominator was N or N^2 or any other variety of factors dependent on N. Of course it can be mathematically shown that the root of N is correct, but for intuition we can look to the law of diminishing returns. If you wanted to cut your level of uncertainty in your average value by half, you would have to do 4 times the number of trials. If you want to be certain of your average to an additional decimal place you have to do 100 times the number of trials! So there is great advantage in performing more than a small number of trials, but eventually the marginal improvement on your uncertainty become relatively small.

Standard Error & the Gaussian Distribution

Let's think about how to get the uncertainty of the average in a different way and return to our experiment to measure the value of g. We could perform 10 trials of the experiment and get an average of those 10 trials. Maybe we get an average of 9.82 m/s^2. We could then do another ten trials and perhaps for the second set we get an average of 9.79 m/s^2. The point is that we would not expect the average of any set of 10 trials to be exactly the same. There would always be some small variation. However, we would expect that variation to be smaller if in each set the number of trials was increased, say to 50. So again we can see the influence of the number of trials on the uncertainty of the average.

What is most important about this method of sampling is that the different averages, (for a constant number of trials), will be randomly distributed. So if we plot the frequency at which we see each particular average value, we produce Gaussian Distributions as shown below.

Figure 5
Experiment to Determine g



The orange line is the same distribution as it was in Figures 3 and 4. We have simply reduced the domain of the *x-axis*. Keeping in mind that the area under the curve represents the probability of finding a value in that range, notice how few individual trials would fall between 9.4 and 10.0. However we see that if we take the average of ten trials, the probability of finding an *average* in that range, (area under the blue curve), increases dramatically. Of course the area under the green curve is higher still and so the probability of getting an average of 20 trials in this range is larger as well.

As these curves are Gaussian, the mathematical properties are the same as the ones we discussed earlier. Here the central value would be the average of the averages taken, which we will assume is the true value of g and the width of the curve is controlled by the standard error. We can also feel free to use the numbers listed in Table 1 above. For the green curve $S = 0.20$ and $N = 20$, so $\bar{S} \approx 0.045$. This means that we could say that our experimental value for g is 9.8 and we can be 95% certain that the true vale of g is within ±0.088 (1.96 * 0.045 ≈ 0.088) of 9.8. Written in proper format our result would be $g = 9.80 \pm 0.09$ m/s^2.

At last we have a way of to present uncertainty for our experiments. We would perform multiple trials and present the average value of those trials as the result with the standard error as basis of uncertainty present in the average value. Note that you need to make clear what confidence level you are presenting your results to. Talk to your TA and see what they prefer. When in doubt specify the confidence percentage along with your result. This is something that we will use when reporting results throughout the semester so please keep this in mind as you perform your lab work.

One Final Adjustment: Population & Sampling

Let's look at a simple example. Say I wanted to know the average height of the students in Monday's lab class. I could certainly measure everyone and calculate an average and standard deviation. Let's say I get the following results (ignoring formatting rules for the moment).

Avg. = 175.2 cm

S = 3.47 cm

N = 24

In this case the standard deviation gives me an idea of what the spread about the average is. The standard error could be calculated, but it would not have much meaning.

Things get complicated when I want to assume the average height of people in this class represents the average height of all students at Emory. If this assumption is to be in any way valid, I would need to put a great deal of effort into making sure that the group of students I use is representative of the Emory student population. This is no easy task. There is a whole field of study devoted to experimental design, and while it will not be something we spend a great deal of time on this semester, if you become a professional scientist, you will certainly have to consider design issues carefully when creating your own experiments. For now just understand

that error can sneak into your results simply by observing only a sample of all possible data.

Now in a physics setting we are attempting to measure physical parameters and therefore we can never measure the full population of values. People could measure values of the speed of light for a billion years and yet they could still keep measuring; the population size is infinite. Therefore it will always be appropriate to consider the measurements we do as a sample of that infinite population.

It can be shown that if we assume that what we are measuring is a random variable, and that each trial is independent of each other one, that increase in error due to taking a sample is dealt with by modifying the standard deviation formula.

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

We will use the greek letter sigma to represent the unbiased standard deviation of a sample. We say unbiased because if the denominator was simply *N* as it was before, our estimation of the standard deviation would be too low. By decreasing the denominator, the estimation of error goes up and this corresponds to the increase in random error introduced into our experiment by sampling

We can now put some meaning to the numerical results we have. For the simple experiment listed on the previous page, we could say…

S = 3.47 cm and this represents the spread of the heights around the class average.

σ *= 3.54 cm and I could be ~68% sure that any one random Emory student falls this far above or below the average of 175.2 cm*

$\bar{\sigma}$ *= 0.724 cm and I could be ~68% sure that the actual average height of Emory students lies this far from the measured sample average of 175.2 cm*

So in the end results can be summarized as such. There is some true value that we are trying to measure. Our best guess at that true value is the average of our sample data set and the unbiased standard deviation will give us our uncertainty in our average value. Keep in mind that you need Table 1 (or a similar table) to convert your errors into uncertainty. Also note that once the standard deviation is corrected for sampling bias, nothing changes about the standard error formula. The unbiased standard error is simply,

$$\bar{\sigma} = \frac{\sigma}{\sqrt{N}}$$

One last word about calculating these values in Excel. There are three standard deviation formulas in the program and here is how they should be used.

=stdev() or =stdev.s(): Both of these functions give the unbiased standard deviation of a sample. This is what we will use the majority of the time.

=stdev.p: This gives the standard deviation of the population. This is the formula from several pages ago when we first wrote an expression for standard deviation, and it should only be used if your measurements encompass the entire population, not just a sample of it.