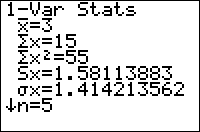
**Data Analysis How-To Guide**

This step-by-step guide will help you know how to generate descriptive statistics and generate graphs in Microsoft Excel (available on all of our computers), as well as calculate descriptive statistics and perform statistical tests on your TI-84Plus calculator. Before you begin, you must know what kind of tests you are trying to do. Computers are good, but they will only do what you tell them to do. Here are a few tips:

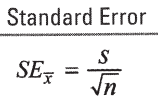
* A bar graph is used to represent categorical data.
* A line graph or scatterplot with a trendline is used to represent continuous data.
* In both cases, the independent variable goes on the X axis and the dependent variable goes on the Y axis. And in both cases it is extremely rare that raw data is plotted on a graph. Typically the mean of multiple trials is plotted with an indication of calculated error.
* Descriptive statistics include the mean (average value), standard deviation (describes how spread out the data are), and standard error of the mean (tells how precisely you know the true mean of the population. It takes into account the standard deviation as well as the sample size).
* A Chi-square Goodness of Fit test examines the relationship of data to predicted results. This is only used when you have a scientific and/or mathematical basis to determine what *should* have happened if your null hypothesis were true (e.g, genetic tests).
* A t-test compares two conditions, often a control against an experimental condition. (e.g., transpiration rate in humid vs. dry conditions; enzymatic reaction rate at pH 3 vs pH 5)
* Last keep in mind what *trials* are: they are **exact** replicas of the experiment. They are used to reduce error. If you’ve changed anything between two trials, they are no longer trials. Trials are represented by ‘n,’ or sample size.

**Calculating Descriptive Statistics**

Smaller Data Sets:

On your TI-84 Plus, enter your data in a list, by pressing “Stat” and then “Edit” followed by typing your multiple trials in a single list. Calculate your descriptive statistics by “Stat” followed by “Calc” and selecting option 1: 1-Var Stats. Then identify the list where you entered your data and press enter. The calculator will generate a screen that looks like this:

To interpret, you really only need two values from this screen: the sample mean, or , and the sample’s standard deviation, or *Sx*. While you may need to be able to calculate the mean by hand, you do not need to be able to calculate the standard deviation by hand. That’s good news: it’s a pain.

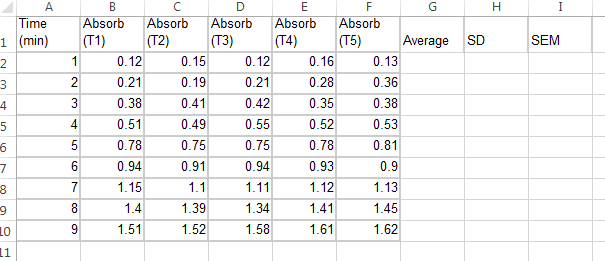


Another descriptive statistic, Standard Error of the Mean (SEM or ), is not represented on this screen. To find SEM, simply divide the standard deviation (Sx) by . This formula is on your formula sheet:

A big warning: recall that all the values you put in the list are trials – repeated *identical* data collection events. This means you did the exact same thing multiple times, or we each did a lab once and the class data were pooled. I cannot stress enough how important this is. Remember, garbage in, garbage out.

Larger Data Sets:

It really is easier to work with larger data sets in MSExcel. Suppose I was interested in measuring the rate of a reaction over time. I performed the reaction 5 times (so 5 trials) and collected data in one minute intervals for 9 minutes during each trial. Going through the process outlined above for descriptive statistics is both tedious and laborious. Instead I enter the data in Excel.



Here you can see the independent variable (time) is in the left column and each of the five trials is listed in their own columns on the right. To ask Excel to calculate descriptive statistics, I title the next columns “Average,” “Standard Deviation” and “SEM.”

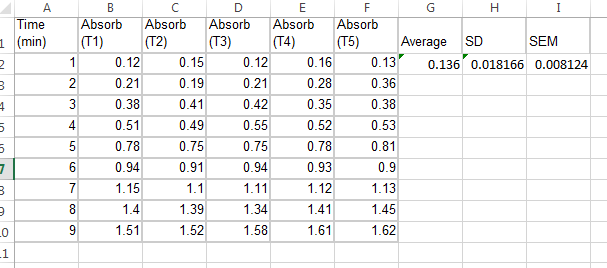
In the row for Time 1 minute (Row 2), I give excel the commands to calculate these descriptive statistics, then drag so that it will autofill. Confused yet? The commands are:

To find an average / mean: =AVERAGE(B2:F2)

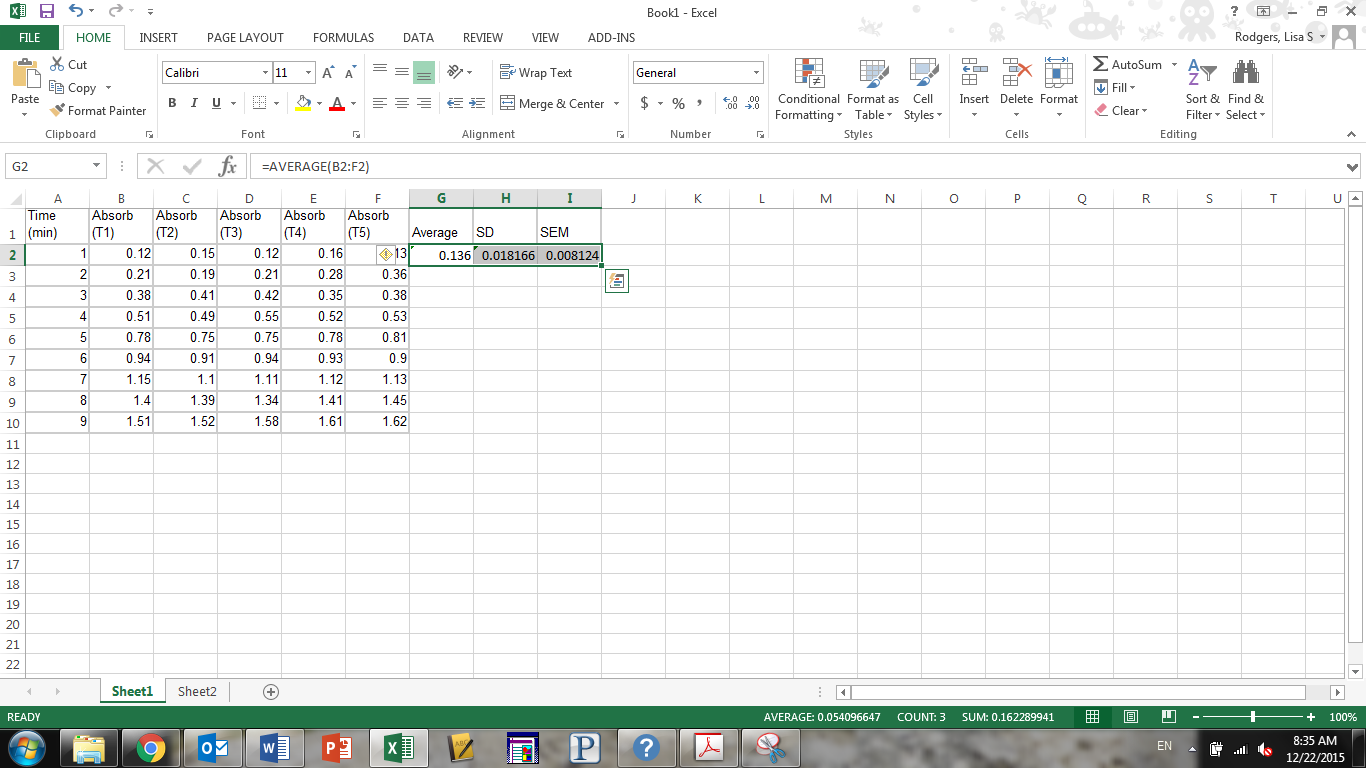
To find the standard deviation: =STDEV.S(B2:F2)

To find the SEM: =H2/sqrt(5)

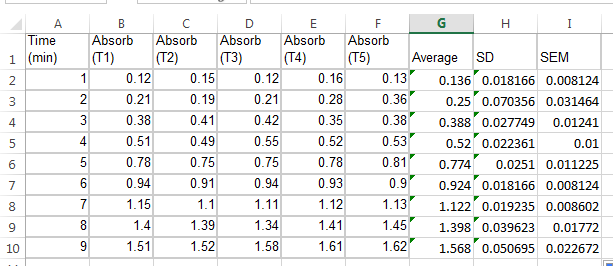
In each of these cases, the B2 and F2 represent the range the data is in. Adjust accordingly. When calculating the SEM, the H2 references the cell the standard deviation is in and the (5) references ‘n’. Now your spreadsheet should look like this:



Now, highlight the three cells where you typed the commands, and place your cursor over the green box in the bottom right hand corner of the highlighted area. Then drag down all the way to minute 9 to autofill the commands.



Your spreadsheet now looks like this:

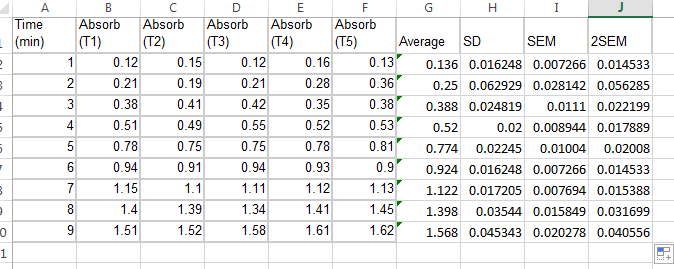


Congrats! You have just computed all of your descriptive statistics!

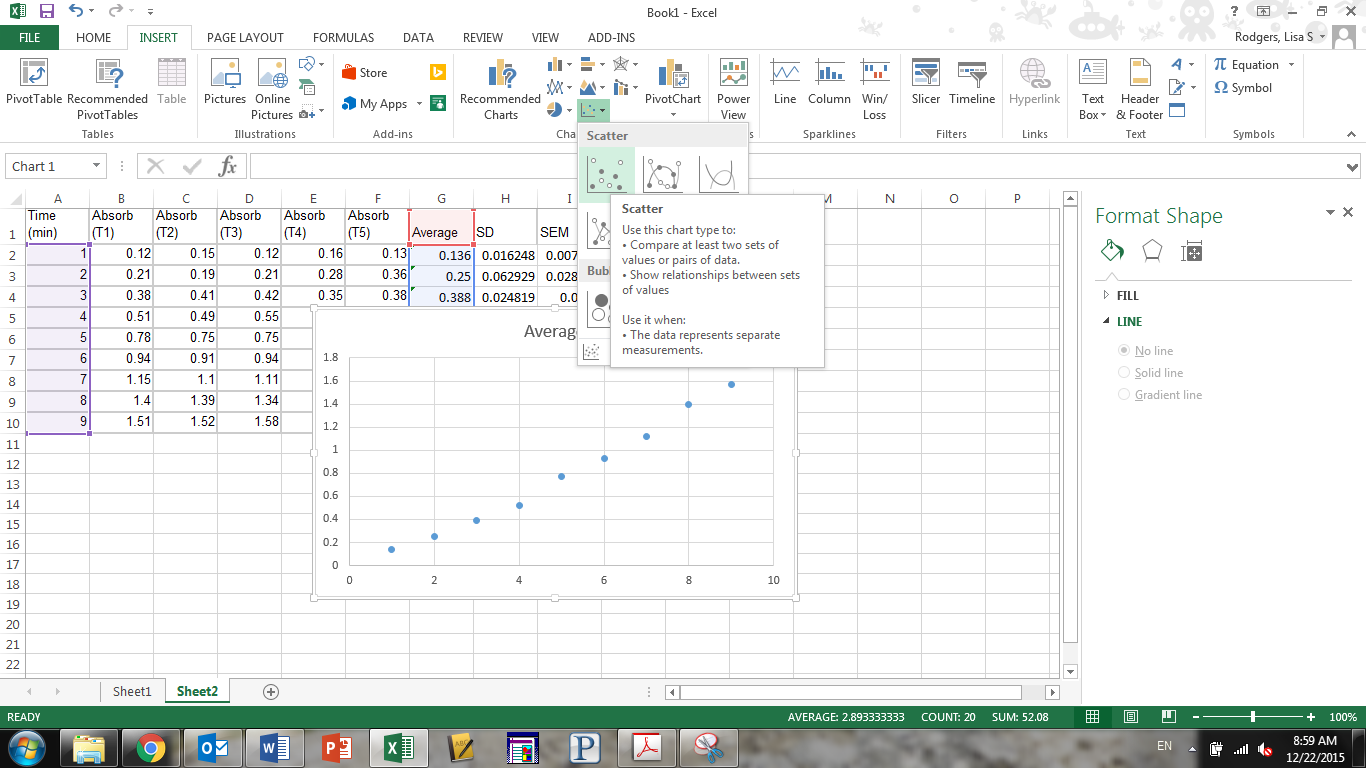
**Creating a Graph**

For most of your graphing work, you’ll just create graphs by hand in your lab notebook. However, for projects and presentations, you’ll need something more professionally assembled. Communicating what you want with Excel can be a pain, but it really is the easiest way once you know what you’re doing. Remember, almost never do you represent raw data graphically. Rather, represent a mean with error bars (typically ±2SEM, a 95% confidence interval) on appropriately labeled axes. We’ll continue with the above example, creating a line graph since it is continuous data.

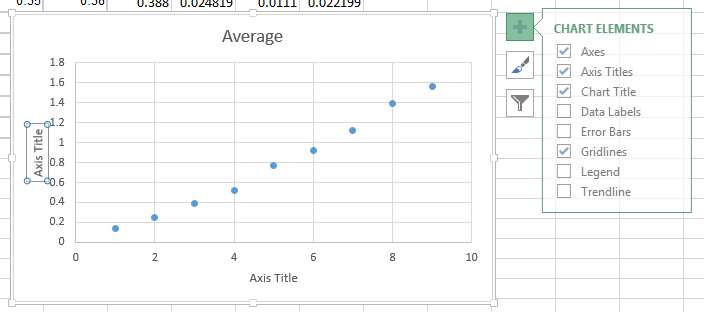
First, create another column in your data table that calculates 2(SEM) so that Excel knows how large to draw error bars. Just as you did before, title the column, and then give the proper command in row 2 (Time 1 minute). The command is “=I2\*2” and then drag down to the bottom of the table to fill in your 2SEM values. Your table should look like this:



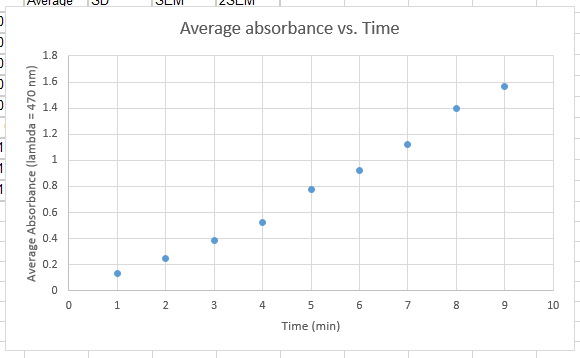
We’re now ready to make a graph. Highlight the cells that contain the values of your independent variable. Press and hold the Ctrl key on your keyboard and then select the AVERAGE of your dependent variable. Then click the “Insert” tab, click on the tiny scatter graph (NOT the line graph!) and choose the upper left hand scatter graph. When you hover over it, it will preview like you see here. You’ll notice it is missing lots of parts, but you’ll fix it up next.

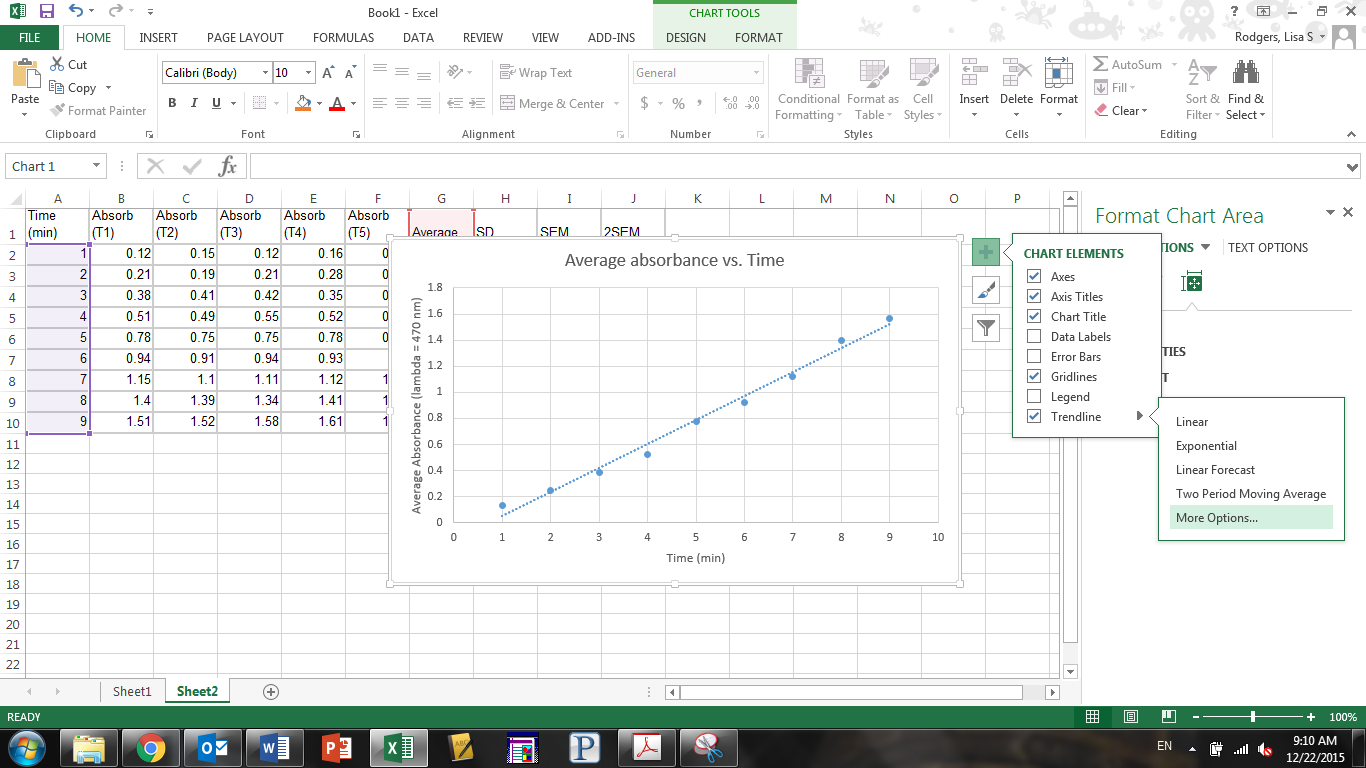


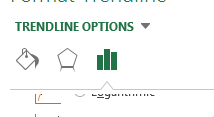
Now you have an ugly graph! First things first, label the axes by clicking on the plus in the upper right hand corner of the graph and selecting “Axis Titles”:



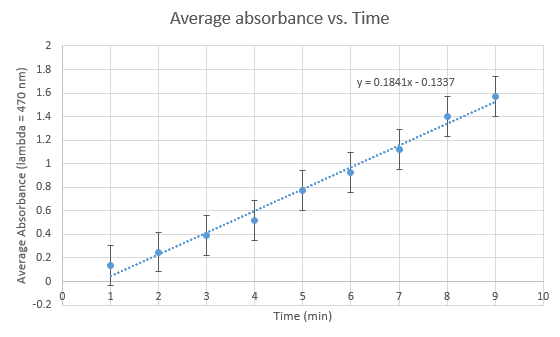
Now click and type in the labels and units in the new text boxes. While you’re at it, fix the title. I really don’t like that the grid on the x-axis goes 0, 2, 4, 6 and so forth. I’d rather it just count by ones. The easiest way to change this is to just make the graph bigger (click the bottom left-hand corner and drag outward) and let it auto-correct. Now you have this:

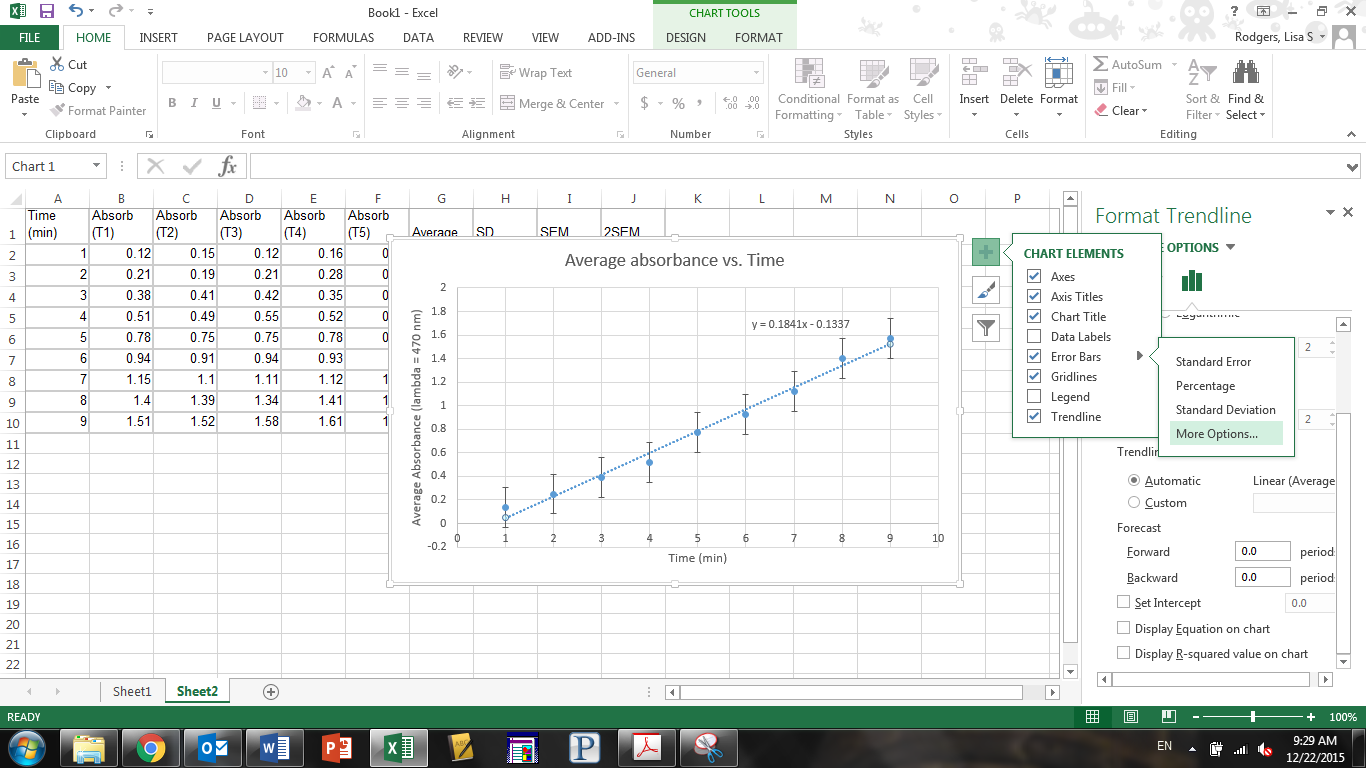


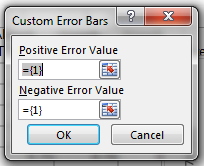
We still need to add a trendline (AKA Line of best fit) as well as error bars. First the trendline. Go back to the + on the upper right corner of the graph and check the “trendline” box followed by the arrow to the right. Then choose more options. 

In the pane on the right, scroll down and choose “display equation on graph” under the third pane of options (indicated by the mini bar graph up top next to the pentagon). If you don’t like where the equation appears, you can click it and drag it wherever you want it.

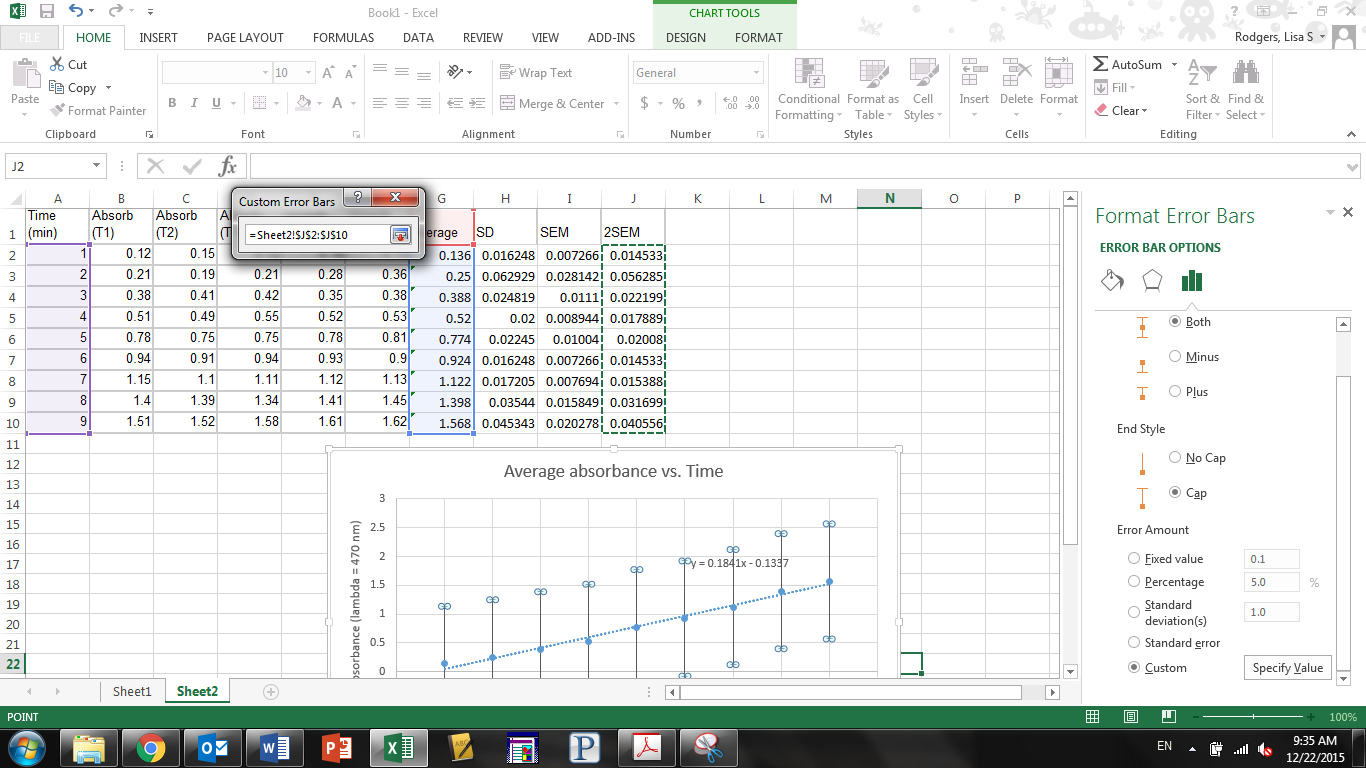
Finally, we’re ready for error bars. Once again, click the + on the upper right of the graph, now choose Error Bars. *You will get error bars in both the X and Y direction. We’ll fix it!* Click on any one of the horizontal error bars, and press delete on your keyboard. All of the horizontal bars will disappear. Now it looks like this:



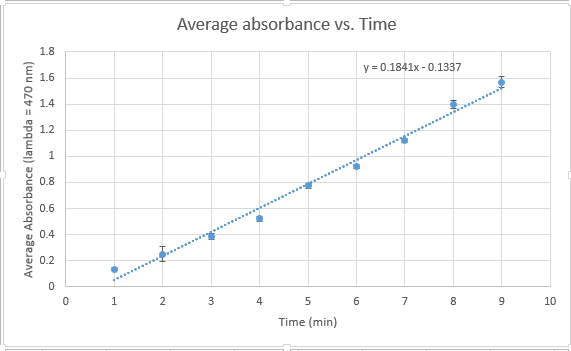
But you are not finished yet. Excel defaults to the error bars being calculated arbitrarily. That’s not going to work, so we need to tell it how large the error bars should actually be: the 2SEM column we calculated earlier. First, click and drag your graph down so that you can see the entire 2SEM column. So go back to the + next to the graph, hover over Error Bars, then hover over the arrow, then select “More Options.”

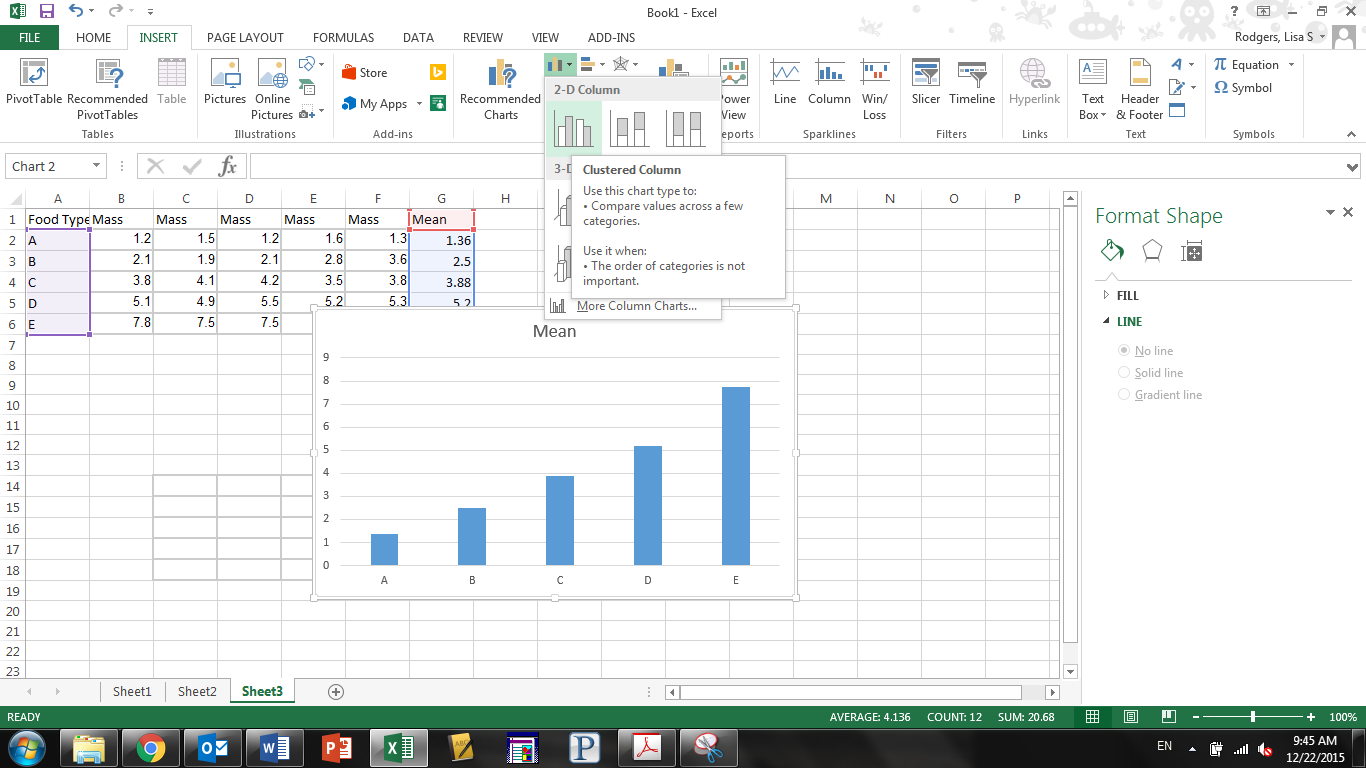
Ensure you’re on the third pane of options and scroll to the bottom. Select “Custom” and click “Specify Values.” This funny little box will appear:

We will now select the 2SEM column we calculated earlier for both the positive and the negative error values. This is achieved by clicking on the red arrow in the box (the box will collapse), then highlighting the column of 2SEM values. There will be a dancing green box around the highlighted values. Click the red arrow again to expand the box, and repeat the procedure for the negative error values.



When you are finished, click “Ok” and watch your error bars take the appropriate size! You have just finished creating a line graph in Excel with all the proper parts. Woot woot!

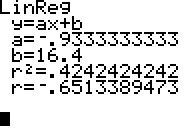


Bar Graphs:

Of course, sometimes you’ll have categorical data where a bar graph is more appropriate. You’ll follow the exact same procedure as above except you will not add a trendline, and when choosing your graph you’ll choose “Clustered Column.” Again, be sure to not represent individual data points. Represent averages with error bars.

Finding a best-fit line using a TI-84 Plus:

Sometimes you’ll be working with individual data points and need to find the Line of Best Fit for the line. Usually this means you’re working on a lab in your lab notebook and you need to find the rate of reaction, or something similar. You may not just connect the first data point with the last data point, or connect the last data point to (0,0) – both techniques your teacher sees regularly and causes her to furrow her brow and groan. You don’t need to work in MSExcel to get the formula for a best fit line. To accomplish this on your calculator:

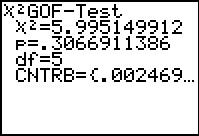
1. Enter the data into lists in your calculator as the coordinate pairs you have graphed. Enter your list of X values in L1 and Y values in L2.
2. Go to “Stat” then “Calc” and choose 4: LinReg(ax+b).
3. Select the lists where you entered your variables, separated by a comma. Looks like this: LinReg(ax+b) L1, L2
4. Out will pop a screen such as this one:
   1. Note that depending on how current your calculator is, you may not get an “r2“ or “r” value. We don’t use them anyway, they indicate correlation and prediction. If you are taking AP Stats, though, you probably will use them.
5. Now you have the slope (a) and y intercept (b) and you can plot your line on the graph you drew in your lab notebook!

**Statistical Tests on your TI-84 Plus**

You can program Excel to do these tests for you, but it’s a huge pain. Just use your calculator. Before calculating any statistical test you need to write your null and alternate hypotheses! Be careful about choosing which test is appropriate. In no case are both appropriate. Think: am I comparing my data to predicted values that are the result of a scientific or mathematical principle? Then use a Chi-square test. Or, am I comparing two data sets I collected against each other? Then use a t-test.

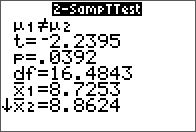
Performing a Chi-square Goodness of Fit test:

Remember, you must be able to do this by hand on your AP Exam, so I’ll ask you to show your work from time to time just to make sure you can do it. ☺

1. Enter your Observed/Actual data collected in List one.
2. Enter your Expected data in List two.
3. Go to “Stat,” then over to “Tests”, then down to option D:*X*2GOF-Test. (Note: the *X*2-Test, option C, is NOT THE SAME. It’s a matrix thing you get to do in AP Stats.)
4. It will default to Observed: L1 and Expected: L2. As long as that’s where you put your data, don’t change those settings.
5. You must enter the degrees of freedom. Remember, the number of distinct outcomes (categories) minus 1. *You may not leave this at the default value and draw quality conclusions.*
6. Then select calculate. You’ll get a screen that looks like this:
7. Note the Chi-square statistic is listed first, followed by the p-value. Recall that when you compute Chi-square by hand you will not get a specific p-value. Rather you compare your Chi-square statistic to the table of critical values on the formula sheet. If your Chi-square statistic is larger than the critical value, reject the null hypothesis. If your Chi-square statistic is lesser than the critical value, you fail to reject the null hypothesis.
8. A final note: the bottom line, called CNTRB, lists the individual contributions of each category to the final Chi-square value. You have to scroll to the right to see the rest of them, but once you press another key on your calculator you can no longer scroll. This is VERY useful tool if you are checking work you completed by hand and you don’t get the same Chi-square value. You can use this CNTRB line to isolate the row where the discrepancy lies and figure out what the error is.

Performing a t-test:

Again, just use your calculator for this task. T-tests REQUIRE multiple trials. You cannot calculate a t-test without multiple trials. For example, if you are comparing the rate of respiration at two different temperatures, you will have established two different water baths, placed multiple organisms – for the sake of this example, let’s say 5 - in each water bath and collect data from a total of 10 organisms. Step by step,

1. Enter the data in your calculator in two lists: List 1 for one temperature; List 2 for the other temperature. In this example, each list consists of the 5 values that were determined from the five individual samples in the same water bath. No mixing and matching.
2. Go to “Stat,” then “Tests” and select option “4:2-sampTTest…”
3. In the screen that appears, verify the following inputs:
   * Inpt: Data
   * List1: L1
   * List2: L2
   * Freq1: 1
   * Freq2: 1
   * µ1:≠µ2
   * Pooled: No
4. Press down to highlight Calculate and press enter.
5. You’ll get a screen similar to this:
6. There’s all sorts of information here. It’s a good idea to verify that your means are accurate (you should have calculated them earlier in the lab as descriptive statistics. Make sure they are equal to 1 and 2 here. If not, you’ve made an error somewhere. Go back and find it before moving forward.)
7. I will ask you to record the t-statistic (here it is -2.2395) and p=value (here it is 0.0392) in your lab notebook.
8. A final note on µ1≠µ2 vs. µ1<µ2 vs. µ1>µ2. The symbol µ represents the true population mean – the statistical parameter you are trying to approximate through your data collection and analysis. If your null hypothesis is “The respiration rate at temperature X is not different from the respiration rate at temperature Y” then µ1≠µ2 is the correct selection. If, however, your null hypothesis is, “The respiration rate at temperature X is lower than the respiration rate at temperature Y” then µ1<µ2 is the correct selection (assuming data for temperature X is in list 1 and data for temperature Y is in list 2).

A final word of caution: Your calculator does literally hundreds of other tests and computations that I don’t describe here. Stick to these few functions for now. Once you have a handle on how it works you can graduate to other functions. Statistics is really cool and gives us a way to defend our conclusions. But it’s all up to interpretation: if you can’t interpret a value in a scientific context, it’s useless to you.