**Using Excel**

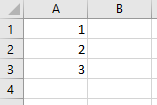
Why use Excel?

* It allows you to organize and manipulate data easily.
* It can create high-quality graphs to include in your manuscript.
* It can be used to perform statistical analyses.

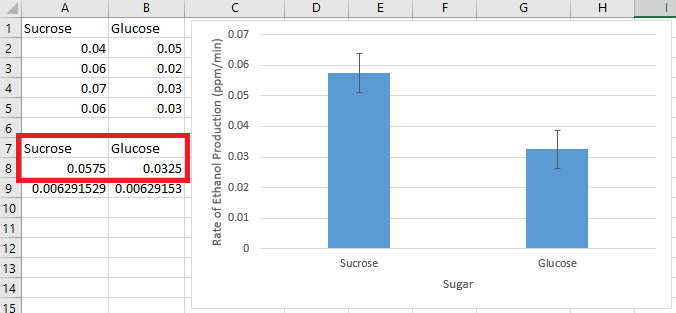
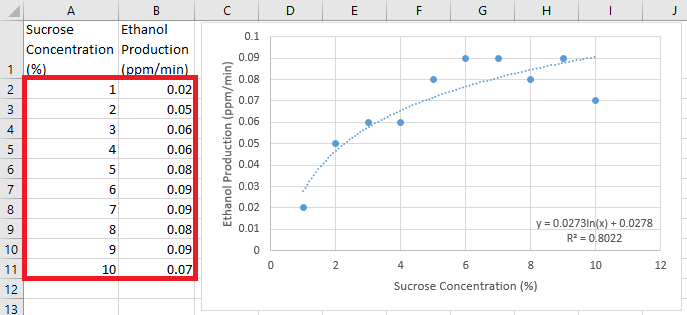
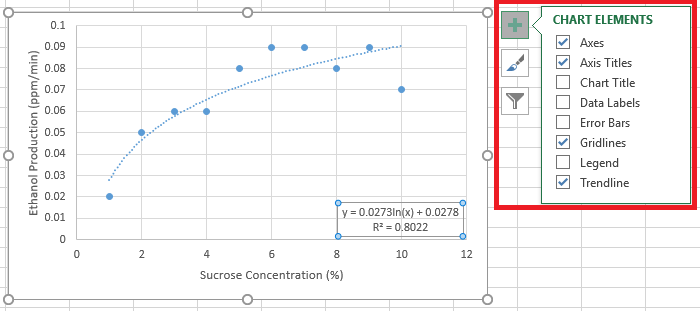
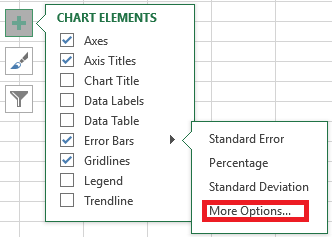
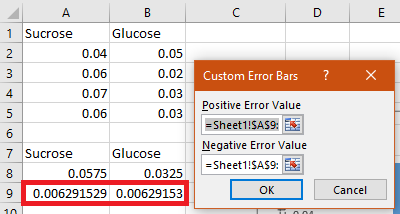
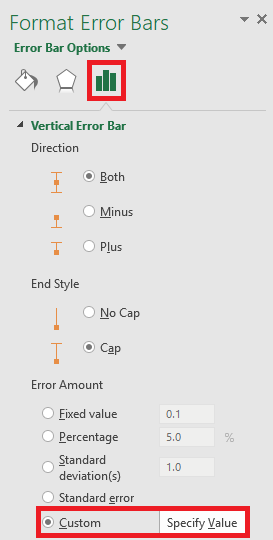
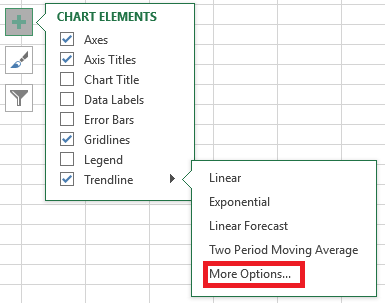
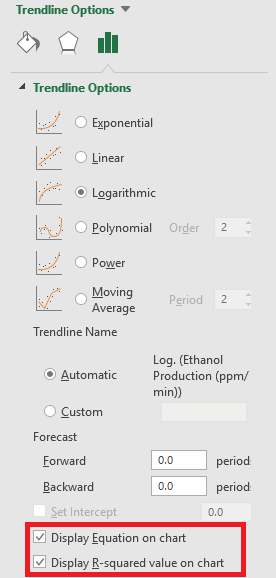
General Use

* Excel is a spreadsheet software; data is entered into a grid, and can be manipulated.
* Some terms:
  + **Cell** - Each square of the spreadsheet is a cell. Each cell has a two-part identifier determined by its row and column.
  + **Row** - All cells in a horizontal line. Each row has a number, starting with 1 at the top.
  + **Column** - All cells in a vertical line. Each column has a letter, starting with A on the left.
  + **Function** - A formula that calculates data in the cell it occupies. Usually, it is used to calculate data using data from other cells.
  + **Workbook** and **Sheet** - The workbook represents the Excel file saved on the computer. Each workbook contains one or more sheets, which can be accessed in the bottom left corner of the Excel window.
* Excel treats numbers and words differently. If a cell contains nothing but numerals and periods, it is a number and can be used in mathematical functions.
  + If you’re recording data, don’t put the units in each cell with the numbers!

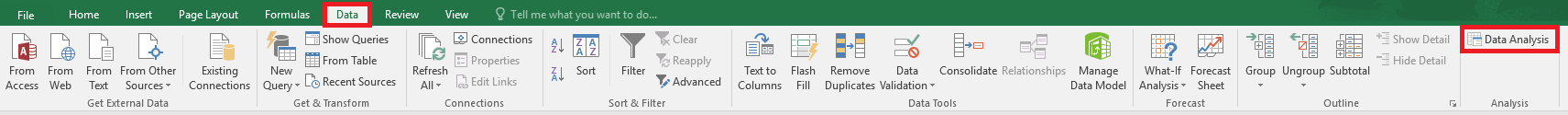
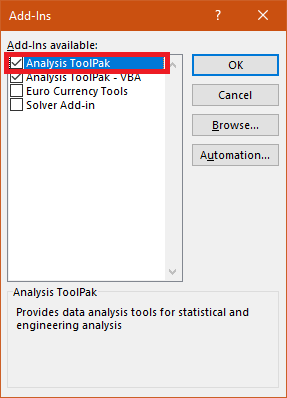
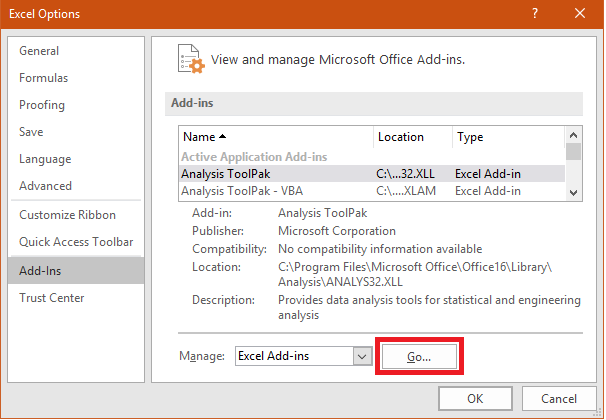
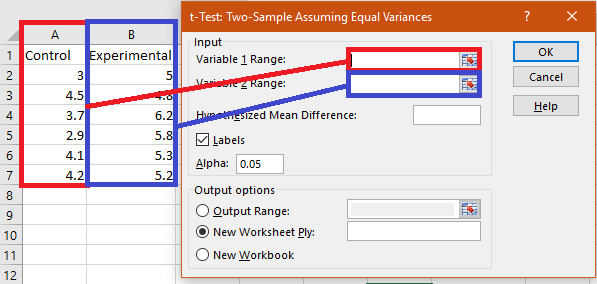
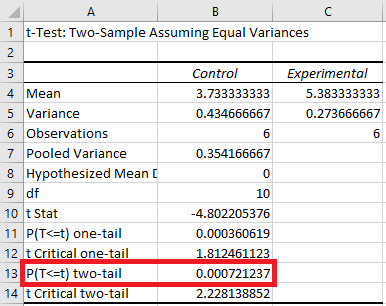
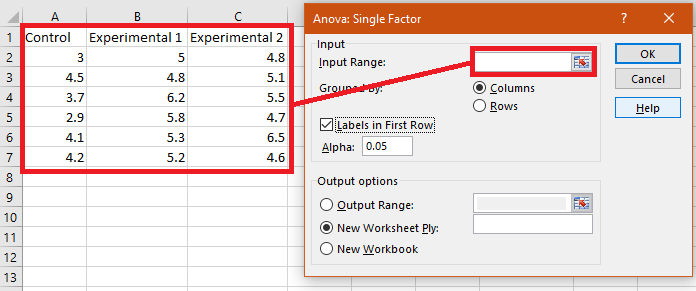
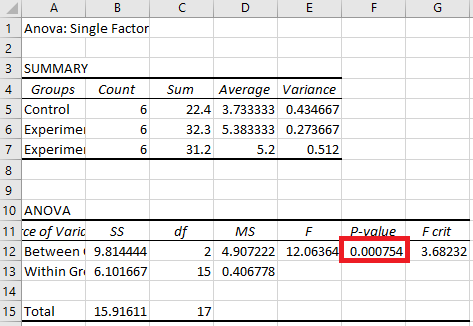
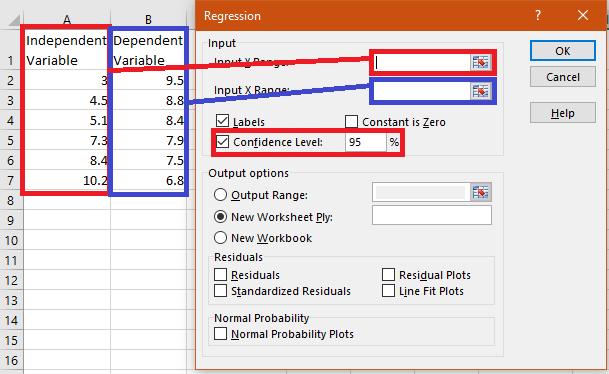
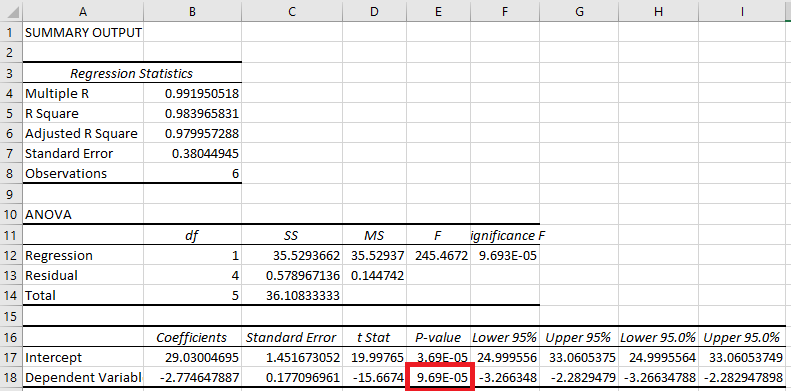
Functions

* Functions always begin with an equals sign (“=”).
* The functions that will be most useful for this course are mathematical.
  + Basic arithmetic:
    - “=2+2” will output “4”
    - “=2-2” will output “0”
    - “=2\*2” will output “4”
    - “=2/2” will output “1”
  + Other functions require that certain inputs be put into parentheses after the formula name.
  + Some functions do simple multi-step arithmetic:
    - “=AVERAGE(2,3,4,5)” will output “3.5”
    - “=SUM(2,3,4,5)” will output “14”
    - “=COUNT(2,3,4,5)” will output “4”
  + Even more complex functions are exist!
    - “=STDEV.S(*values*)” finds the standard deviation for the selected sample.
    - “=SLOPE(*y-values, x-values*)” finds the slope of a line of best fit.
* Functions can use data from other cells.
  + For example, if you have the following data...  
      
    ...the function “=AVERAGE(A1:A3)” will output “2”.
  + If you copy and paste a function, it will use data from the same *relative* position.
    - So if the function “=AVERAGE(A1:A3)” is in cell A4, and you copy that function to cell B4, the new function will be “=AVERAGE(B1:B3)”.
* Functions can be nested within each other. They follow mathematical order of operation, so you can use parentheses to control what order the functions act in.
  + For example, you can use the formula   
    =STDEV.S(A1:A10)/SQRT(COUNT(A1:A10))  
    will calculate the standard error for the range of numbers in cells A1-A10.
  + Note that you do not need to precede each function with an equals sign, just the whole formula.

Graphing

* Before you begin making a graph, you should determine what sort of graph best represents your data.
  + **Categorical** - If you are comparing two or more different treatments, you should use a **bar chart**.
    - For example, you give yeast either glucose or sucrose and test which one causes the yeast to produce the most ethanol.
  + **Continuous** - If you are determining if (and how) your dependent variable changes with a continuous independent variable, you should use an **x-y scatterplot**.
    - For example, you give yeast different concentrations of sucrose and measure ethanol production.
  + Other types of data may require different graph types, but these are the most common types in this course.
* All graphs are made by selecting the appropriate **Charts** option in the **Insert** tab.
* **Bar Graphs** compare the averages of multiple different treatments.
  + In order to create a bar graph, you need to calculate the average value of all the trials in each treatment. Format the data like this:  
      
    By selecting the highlighted data and selecting the **Clustered Column** chart, you will create a simple bar chart. You will need to remove the chart title and add axis titles, and you may also want to include error bars. Details on all of this can be found below.
* **X-Y Scatterplots** show each trial as a point on a plane, with the independent variable on the y-axis and the dependent variable on the x-axis.
  + You should format your data like this:  
      
    By selecting the highlighted data and selecting the **Scatter** chart, you will create a simple x-y scatterplot. You will need to remove the chart title and add axis titles, and you may also want to include a trendline. Details on all of this can be found below.
* For scientific manuscripts, your graphs SHOULD NOT have a chart title, and SHOULD have axis titles.
  + You can modify these things by clicking the + box at the top right of a selected graph:  
    
* When adding error bars to a bar chart, Excel will give you a default option, but it is not correct.
  + First, you must calculate the standard deviation or standard error for each treatment (it’s easiest to calculate this right below the average). Then, on the + menu, click on the arrow next to **Error Bars** and select **More Options…**  
      
    This will open a side menu. Make sure you have selected the bar chart icon, then select **Custom** and click the **Specify Value** button. Now enter your standard error/deviation values into BOTH the positive and negative value ranges.  
      
    
* When adding a trendline to a scatterplot, a linear trend is often sufficient, but Excel offers a few other options. The previous examples shows a logarithmic trendline. You can access these options using the + menu and by selecting **More Options...**  
    
    
  Note that you can also display the equation for the trendline, as well as the R2 value (which represents how close the line is to the data points).

Statistical Analysis in Excel

* Before using statistical analyses, it’s important to understand what they mean!
  + Often, when we see differences in data, we’re not immediately sure if the difference is big enough to be meaningful. There is always the possibility that the difference was caused by random chance!
  + A statistical analysis determines the *confidence* (represented as a percentage) that the differences are just due to chance.
    - This confidence is presented as the **p-value**, which ranges from 1 (100%) to 0 (0%).
    - A *low* p-value (less than 0.05) means that your hypothesis is supported.
* Excel contains tools that make statistical analyses much easier. **Data Analysis** can be found under the **Data** tab, usually at the far right.  
  
  + If it’s not there, you may need to activate it by going to **Options** under the **File** tab. Go to **Add-Ins**, click the “Go…” button at the bottom, check the box next to **Analysis TookPak**, and click “OK”.  
    
  + Clicking on **Data Analysis** will give you a list of options...
* **t-Test: Two-Sample Assuming Equal Variances** compares two different treatments (for example, control vs. experimental) to see if there is a meaningful difference between them.
  + Put all the values from each trial into the inputs as shown:  
    
  + Check **Labels** if you included the title of the columns in your ranges.
  + You should get an output that looks like this:  
      
    The highlighted value is the p-value!
* **ANOVA: Single Factor** compares three or more different treatments (control vs. two different experimental groups) to determine if there is a difference between any two of them.
  + Put all the values from each trial into the inputs as shown:  
    
  + Check **Labels** if you included the title of the columns in your ranges.
  + You should get an output that looks like this:  
      
    The highlighted value is the p-value!
* **Regression** determines if there is a meaningful relationship between two continuous variables.
  + Put all the values from each trial into the inputs as shown, and make sure to check the box next to **Confidence Level**:  
    
  + Check **Labels** if you included the title of the columns in your ranges.
  + You should get an output that looks like this:  
      
    The highlighted value is the p-value!  
    **Be aware that Regression only tests for linear relationships!**