Technical Intro to Scientific Data Analysis with Excel (Excel 2010 version)

Excel (or any spreadsheet program) can help you do calculations quickly and efficiently. But Excel is only a software program—it can only be as smart as the instructions that you give to it. Today you will practice and master the skills necessary to bend Excel to your will and make it do the grunt work. After today, you will ALL be expected to be experts at these skills so take turns and help each other learn. Take notes for the future if you are worried that you will forget. Each version of Excel (as with the other Microsoft Office programs) varies slightly, but familiarity with one version should help you intuitively guess/explore the other versions. Some of you may feel that you are already familiar with Excel; please READ the Technical document anyway! It contains specific scientific norms that you need to learn.

* Entering Data:

In a blank/new Excel document:

To enter information in a cell, click on the cell (e.g., cell A1—column A, row 1) and type the information. When you are finished, press 'Enter.' You will see the information in the cell. To edit the information, click on the cell and type (erases previous entry) or click on the cell and then click on the formula bar to edit (the formula bar is above all the cells but below all the icons for text manipulation (bold, italicize, center in cell, etc.).

Title each column with the quantity's **name and units**. In the cells beneath each column title, enter the data that you have collected by typing the numbers into the cells. When you have finished entering your data, **save the file** by pressing 'Ctrl' + 'C' (name the file and note the saved location for future use) or by selecting the 'File' tab, then the 'Save as' icon. **Save regularly** to avoid losing work. Every time you save the file, it overwrites the previous version.

The **number of decimal places displayed** in the cell can be controlled by icons in the 'Number' menu on the 'Home' tab. Use these icons to give your data uniform appearance.

* Generating Columns of Data Calculated from a Formula:

To generate information from a formula (i.e., to mathematically manipulate cells), click on a cell and **begin with =**. All Excel entries for which you expect a numerical result/output MUST begin with an equals sign, =. The mathematical manipulators are what you would expect: * to multiply, / to divide, + to add, and – to subtract. To exponentiate, type ^ and the power (e.g., 6^2 is 6^2). Let us say you wished to multiply the entry in cell A1 by 60. To do this you would click on the cell where you wish the output to go and then type '=60*A1.' You can type the input cell, A1, on the keyboard, or you can use the mouse to **click on cell** A1 while typing the formula. For more complex mathematical operations, parentheses often become necessary. Excel will color-code the parenthesis to help you see where each set opens and closes. Be VERY careful with your **parentheses**! (Again, Excel is only as smart as the instructions that you give to it!)

(Generating Columns of Data Calculated from a Formula, Cont.)

If you wish the same mathematical formula to be **applied to every cell in a column** (e.g., column F is column A times 60), type the formula into the output cell for the first row. Then click on the output cell and move your mouse over the bottom, right corner of the cell. Your pointer will change shape and become a plus. Press down the left mouse button, **drag down** to the last cell you wish to effect, then release the mouse button. The cells will automatically fill in. By clicking on any of these cells, you can see in the formula bar that the formula has been adjusted to reference the correct input cell (i.e., for the row you are currently in). The same process can be used to copy a formula across a row into multiple columns. If, in a formula, you wish to reference a **specific column or cell (that will NOT change** when the formula is dragged over a column or row), **use the $ symbol**: e.g., $A1 will always be column A, but the row number can change; and A1 will always be column A and row 1—neither can change.

* Generating Graphs:

When creating a graph/plot, Excel will usually plot the first column on the independent/horizontal axis and the second column on the dependent/vertical axis. You should <u>always check</u> that the correct data has appeared on the correct axis. If the data has been entered with the columns in reverse order, this can be fixed after the plot is created. To generate a graph/plot, use the mouse to highlight the data that you wish to graph (if the data is adjacent, just highlight the entire chunk; if the columns are not adjacent, highlight each column separately while holding the 'Ctrl' key). Once the data is highlighted, click on the 'Insert' tab and then the '**Scatter**' icon (on the 'Charts' menu—other types of charts may be appropriate for different data types, but the Scatter plot is the most commonly used). From the sub-types of scatter plot available, choose the one that has **data points but no lines**. A graph will appear—but you are NOT finished yet! The design of the graph can be adjusted using the 'Chart Layout' menu—you will want to be able to **label the individual axes with the quantity graphed and its units** and you will want to be able to **title the graph**. When the graph has been titled (for the 'vs.' format, it is always 'Dependent' vs. 'Independent') and the axes have been properly labeled, you can choose to keep or delete the Legend (not needed if only one data set is plotted, absolutely necessary for multiple data sets). The location of the chart can be changed by: left-clicking on the chart (the upper right corner works well), then right clicking and choosing 'Move Chart' from the menu. It is often best to make the chart a 'New Sheet' (so that it doesn't block your view of the data)—give it an appropriate title and click 'Okay.' You chart will become its own sheet—and your data will disappear! But the data is <u>not</u> gone—to find it, use the tabs on the <u>bottom</u> of the Excel window ('Sheet 1' is usually the data; now might be a good time to rename it 'Data': right click, select 'Rename' and name it).

Before you finish, **check that the correct data has been plotted on the correct axis and that the axis labels match the quantities plotted**. If the original columns were in reverse order, this can be adjusted by left-clicking on a data point (to select the data), and then right-clicking and choosing 'Select Data'. Click on the Legend Entry you wish to adjust, and then click 'Edit.' Highlight the appropriate Series X and Series Y data and then click 'Okay.'

* Adding a Best Fit ("Trendline") to a Graph:

To add a line of fit (called a "Trendline" in Excel) to a graph, left-click on a data point to select the data, then right-click and choose 'Add Trendline.' Choose the appropriate Trend/Regression type, and tick the boxes for 'Display Equation on chart' and 'Display R-squared value on chart'. As a general rule, do not 'Set Intercept' to a specific number (i.e., do not force the origin to be part of the trendline). Do you know what the R-squared value really means? If not, you should look it up!! Look at the equation for the trendline and think about it. **What does the slope mean? What does the y-intercept mean? What does this R-squared value mean?**

* Adding Error Bars to the Data Points on a Graph:

Error bars represent the uncertainty of the measurement of the data. Do you have uncertainty? Do you have uncertainty in both quantities (horizontal and vertical)? Is the uncertainty the same for all data points, or does it change with the value of the quantity plotted? Wherever you have uncertainty, you must have error bars! Once you have determined the amount of uncertainty, you can add error bars by clicking on the 'Layout' tab (inside the 'Chart Tools' tab) and clicking on the 'Error Bars' icon. To get both horizontal and vertical error bars, a good choice is the 'Error Bars with Standard Error,' and then click on the vertical and horizontal bars (individually) and right click to 'Format Error Bars.' Apply the error bars according to the decisions that you have made about your uncertainty. To remove either vertical or horizontal error bars, left-click on the bars, then right-click and choose 'Delete.' (Future Lab Skill Goal: How do we propagate uncertainty into calculated quantities?)

* Other cool tools:

As a spreadsheet program, Excel can do a lot of useful statistical analysis and mathematical manipulation. Some functions you will find useful include: Sum, Average, and Standard Deviation.

To Sum elements, click on the cell in which you wish the result to appear, then enter '=sum.' Immediately, Excel will list possible formula options that begin with 'sum'. For a simple summation of elements, you would choose the **SUM** function, which you have already typed, but you should note the other options available. Open parentheses following your sum, so that you have now types '=sum(' and then highlight the cells to be summed. Close your parentheses and press 'Enter.' (It would look like this '=sum(A1:A5)' for a sum of the cells in rows 1 through 5 of column A.)

To Average elements, click on the cell in which you wish the result to appear, and then enter '=average(.' Highlight the cells you wish to average and then close your parentheses and press 'Enter.' Note that Excel will offer you a list of average functions—the most commonly used option is the **AVERAGE** function. (It would look like this '=average(A1:A5)' for the average of the cells in rows 1 through 5 of column A.)

To perform a Standard Deviation on a set of elements, click on the cell in which you wish the result to appear, and then enter '=stdev(.' Highlight the cells you wish to perform a standard deviation upon and then close your parentheses and press 'Enter.' Note that Excel will offer you a list of standard deviation functions—the most commonly used option is the **STDEV** function. (It

(Other Cool Tools, Cont.)

would look like this '=stdev(A1:A5)' for the standard deviation of the cells in rows 1 through 5 of column A.)

To perform any of these functions on disconnected sets of cells (e.g., cells in rows 1 through 13 and then 15 through 17), open your parentheses, highlight the first contiguous group, enter a comma, then highlight the next contiguous group, enter a comma, and so forth, continuing to the last group of numbers and ending with a closing parenthesis and the 'Enter' key. (As an example, '=average(A1:A13,A15:A17)'.)

Excel has many other functions available, including trigonometric and logarithmic functions. Investigate what is available by clicking on the 'Formulas' tab and looking through the 'Function Library,' especially the 'Math & Trig' and 'More Functions' menus.

~~~~~~

Notes: