Excel Tips and FAQs - MS 2010

Remember to save frequently!

Part I. Managing and Summarizing Data

NOTE IN EXCEL 2010, THERE ARE A NUMBER OF WAYS TO DO THE CORRECT THING!

FAQ1: How do I sort my data?

Sorting provides a way to rearrange the rows so that all the data for a particular group, like males or females, are put into a continuous block of rows.

1. Highlight every column that is part of the data set.
    - *Beware: If you don't highlight all the columns, you will sort some columns but not others, and the data in your rows will turn into a jumbled mess!
2. Go to the **HOME** tab and select **Sort and Filter button** at far right. Click on **Custom Sort** in the pull down menu.
3. **Choose the variable by which you want to sort by** using the down arrow in the sort by column box.
    - Be sure the "My Data Has Headers" box is highlighted at the top right. If you don't, Excel will sort these cells with the rest of your data
    - For example, if "Gender" is in column A and you want to separate the data by Gender, choose the variable "Gender" for sorting. Thus A-Z is appropriate ordering.
    - If you want to sort by a quantitative character (e.g. shell length) then Excel auto turns "sort on" to smallest to largest.
4. If you want to sort by two or more columns then click add level from the Custom Sort window and then follow the procedures above.

FAQ2: How do I write a function in Excel?

Functions are used to do calculations on numbers in cells.

- An **average** is a common function that we will employ. The arithmetic mean is a *summary statistic* that provides one estimate of what the average member of a sample group is like. The mean is useful for making comparisons between groups.
- The **standard deviation** is a *summary statistic* that provides one measure of *spread* or *variability* in the data. To compare two groups, one must know not only the mean but also the amount of variability in the data. (After all, the means for two groups will likely be at least a little bit different—the question is, are they different enough to conclude that there is a real difference between the groups?). Means are usually reported along with standard deviations or standard errors (see below).

How to calculate an average:

In MS Excel 2010 – there are multiple approaches to getting the average.

1) Functions may be found under the FORMULAS tab under the MORE FUNCTIONS icon (orange books). The first choice in this menu is "statistical" and this is where we will find everything we need.
2) In Excel 2010, there are autofills that will help along the way. For example as soon as you type an equals sign = and 'a' into an empty cell you will see a menu of options. Click on Average = AVERAGE(number1, number2). Then you will select the column(s) of numbers that you'd like to include, with your cursor.
3) Another approach is to click on the ∑ Autosum from the HOME toolbar at the far right. From the pull-down menu from Autosum choose Average (by a double click). Then use your cursor to select the range of numbers that you'd like the average for. Then hit enter and your average will appear in the cell.

How to calculate standard deviation

To conduct a Standard Deviation in version 2010 choose the **STDEV.S** for our purposes in this course.

1) The STDEV.S functions may be found under the FORMULAS tab under the MORE FUNCTIONS icon (orange books). Search for "STDEV.S" in the "statistical" menu.
2) Using autofill in Excel 2010, type = and 'st' into an empty cell you will see a menu of options. Click on = STDEV.S(number1, number2). Then you will select the column(s) of numbers that you'd like to include, with your cursor.

Calculating averages and standard deviation in Excel version 2001.

FYI, in Excel versions 2007 and earlier (such as you may have on a home computer), a function is typed into a blank cell in the following form:

= FUNCTION (*argument 1, argument 2, ….*)

You will use functions like AVERAGE and STDEV to calculate means and standard deviations. The "arguments" used by each function, located inside the parentheses, are the columns of data that you would like to include.

To get the average of a column of data one form is:

= AVERAGE (B2:B9)

Once you have the =AVERAGE in the cell, position the cursor inside the empty parentheses, and then use the cursor to select and highlight the block of cells that make up the argument. This action will put the correct argument between the parentheses.

Calculating averages and standard deviation in Excel version 2001.

Note that you can also have Excel do mathematical (e.g. arithmetic across columns) formulas. For example, if mass is in cell B2 and height is in cell C2, you can enter this formula into a blank cell in the next column (for example, D2) to calculate their ratio:

= B2/C2

FAQ3: How can I repeat the same function without having to retype it multiple times?

Sometimes you will need to make the same calculation on every row of your spreadsheet. For example, you might want to calculate the ratio of mass to height (see example in FAQ2), and then use this ratio as a new variable. Start by entering the formula in a blank cell on the same row as the data. Then you can copy and paste the function into all the cells in the same column on the other rows where you want to make the same calculation. This step can save a great deal of time if you are making the same calculations for every row in a database.

You may also do filling from a Fill tab at the right hand size of the HOME tab.

[To understand why filling works, you must understand the difference between a *relative reference* and an *absolute reference*. In the example from FAQ2, the formula in D2 refers to cells B2 and C2. However, the formula is actually telling Excel, "take the number in the cell two columns to the left and divide it by the number in the cell one column to the left." This is a relative reference—it does not refer absolutely to B2 and C2, but rather to cells that are in the position of B2 and C2 *relative to* D2. So, when you copy and paste your formula from D2 to D3, the formula will do the calculation on cells in the same *relative* position (namely, B3 and C3).]

You should always check to be sure you are getting the result you intended. If necessary, you can correct the reference in the function.

Part II. Making graphs

Tip: Before graphing anything, first figure out what your graph will look like.

- Start by visualizing and making a general sketch, asking yourself the following questions.
  - Will a bar graph, scatterplot, or some other type of graph be most appropriate for the relationship I want to show?
  - How will I label the axes be labeled, and what will need to go in a figure legend?
  - Will I need error bars, and what will they show?
  - Will black-and-white be sufficient, or will color be useful for conveying information?
- Figure out whether your graph will require raw data or only means and standard deviations.
  - If you will be displaying means and standard deviations (for a bar graph, for example), you'll need to organize these descriptive statistics into a mini-table in your spreadsheet.

1. HOW TO MAKE A SCATTERPLOT

- Highlight the cells that include the numbers that you want to plot, *including* the names of groups or variables.
- From the Insert tab, click directly on the type of graph that you would like to make. We'll focus on Column and Scatter type graphs. In scatter you want the **one without lines** (First choice). This hand out will show later you how to insert best fit lines.
- From Chart Layout – choose the example without Gridlines (those horizontal lines throughout the graph). So your graph primarily has a white background.
- In the new green tab, chart-tools, we can examine layout options.
- For example one option is axes titles. Where you will want to choose a title below the horizontal axis and to the left (or side) of the vertical axis. Excel will insert generic axis titles, you will need to click on these and edit them to the biologically appropriate title, with units (e.g. g, mm etc).
- Please delete any Excel inserted legend when there is only one set of numbers included. Legends may be appropriate when there are two sets of numbers (e.g. color codes of two separate species).
- You may wish to adjust the range of your axes so your data are better centered. To do this, choose the axes tab and for example you may pick the 'primary-vertical axis", and then select "more primary vertical axis" options, and in this menu you can insert a new minimum and maximum of the axis.
- To add a trendline, under the 'layout' tab, there is a trendline option. Choose more trendline options. Make sure the trend type is set to linear and you can check the appropriate boxes to display the equation and the r^2 value on the chart.
- If you want to add a second set of data (e.g. a second species with its own trendline instead of grouping the two species), then under the 'design tab' choose 'select data'. On the left where it says legend entries, click on the add button. Using the new pop up

window, selected the appropriate x and y values from your data sheet. In this case a legend may be appropriate, and you will need to create a second trendline following the directions above for the new set of data.

- Otherwise, choose colors and shapes of data points as you wish. Thickening lines is important for printing or projecting. Make sure fonts are large enough to be readable and bold.

2. HOW TO MAKE A BAR GRAPH

- Make a little table of means and standard deviations, with the names of your categories.
- Select the titles and averages of your categories.
- From the Insert tab, click directly on the type of graph that you would like to make. This time you'd like the 2D column (the leftmost option is probably best). You will need to remove the gridlines and legend from this. You will also need to name the Y and X axes following the same procedure as the scatter plot.
- To add the Y error bars, it is under the "chart tools" "layout" tab error bars option. Select more error bars options. Choose custom, and display direction plus. Under custom specify value. From the pop-up menu, select positive error value and then select from the Excel Table the standard deviations that correspond to your two species (the dark grey cells below). **IMPORTANT, for the negative error value enter 0.** Leaving this blank will result in error bars of one, which is not what you are looking to report.
- Otherwise, choose bar colors as you wish. Thickening lines is always good for printing. Make sure fonts are large enough to be readable.
- If needed, change the title text by clicking the text box and adding text. Highlighting text will also reveal a menu for formatting title text.

Save your work frequently!

Tip: How to make your graph look more professional?

Compare your original sketch with the graph you've made in Excel. It may be that the Excel graph is not the clearest. The labels may not be quite right, the colors may be unappealing, the font may be too small or too large, the legend may be unnecessary, and the gridlines may be annoying.

- Right click on axes to change the range for this axis, increase the font size, etc.
- You can also right click on axes and format them to thicken the axis lines
- Right click on bars or points to change colors, add or modify error bars, add a trendline, etc.
- Right click on the whole plot to change colors or to return to chart options where you can remove gridlines or legends, change labels, etc.

- Explore these options until you have an informative, attractive, and easy-to-read graph.
- Think about the data to ink ratio
- Some suggestion for of a clean look of an Excel Graph (Yes, Excel has a mind of its own – you'll need to wrangle it to get it to look like you want it to look)
    - X Y scatterplot graph:
        - *Thick lines*
        - *Labeled axes*
        - *Large readable font*
        - *Simple shapes*
        - *No gridlines or extraneous color*
        - *Simple yet thick trendline pattern*
    - Bar Graph
        - *Thick lines*
        - *Labeled axes*
        - *Large readable font*
        - *Simple colors/hatching*
        - *No gridlines or extraneous color*
        - *Thick error bars*

Part III: HOW TO PERFORM STATISTICAL TESTS IN EXCEL 2010

- Below are instructions for doing a **Correlation Analysis**, **and t-test** using EXCEL 2010.

1. How to perform a Correlation Analysis using Excel

Correlation analysis provides a way to test the hypothesis of a relationship between two continuous variables.

5. Click in an empty cell where you want to calculate the test statistic, ***r***.
6. Click on the formula's tab, then click on more functions (orange books), use the top option Statistical and select CORREL.

   In the array 1 box, click on the little box with a red arrow and then select the column of numbers for your first quantitative trait. It will automatically switch you to the next array (array 2). Array = your column of data for a single trait.

   =CORREL(**array1**,**array2**)

   array1 = the range of cells that contain values for one continuous variable
   array2 = the range of cells that contain values for the second continuous variable

7. Press OK.
8. You will see the test statistic in the cell.
9. Re-select the cell that contains *r*, select **Copy** (or hit Control C) on the **HOME** tab select **Paste on the left hand side of the** menu (with the clipboard), with the tiny black downward facing arrow choose **Paste Values** the first choice (the 123 clipboard). This step pastes the number instead of the function back into the same cell. Why might this be important? If you make a little table in excel on the same sheet of all your statistical test results.
10. A positive value of *r* indicates a positive correlation, while a negative value indicates a negative correlation. To see if the correlation is significant, compare the absolute value of your *r* to the critical value in **Table 1**. To find the critical value you will need the **degrees of freedom** and **alpha** (see handout 1.3, part E).
11. If *r* is greater than the critical value for your chosen **alpha**, you can reject the null hypothesis in favor of the alternative. To report the ***P-value*** for your test, determine where ***r*** falls relative to the other critical values on the same row for your *df* (the possible choices are $P > 0.05$, $P < 0.05$, $P < 0.01$, $P < 0.001$ it may also be between these columns such that $0.01<P>0.001$).

*Beware! If you don't use **Copy** and **Paste Special…Values**, the formula you typed could produce a different number if you sort your data!

2. How to perform a t-test in Excel

The t-test provides a way to test for differences in the mean value between two groups.

1. Check the standard deviations of the measurements from the two groups that you want to compare. If they are similar, use the formula for equal variances below. If they are very different, use the formula for unequal variances. Rule of thumb when in doubt pick unequal.
2. Click in an empty cell where you want to calculate the **p-value**.
3. Under the Formula tab, click the More Functions (orange books) draw down menu, and choose T.TEST.
4. Using the window that pops open, select your columns of numbers for the trait of the two species you want to compare for the **Array1** and **Array2 (**e.g. length of ponderous arc for Array1 and length of incongruous arc for Array2). For the **Tails line** – insert the number 2, as we will be conducting a 2-tail test (what is the difference between a two tail and a one tail test? If apriori you don't know if one is bigger than the other, then you will be testing both the bigger and small hypothesis. Thus the two tail test is appropriate for the work we're doing in this class). In the **Type** line use 2 if your samples have relatively equal standard deviations, use 3 if your samples have quite different standard deviations.
5. Hit enter and the P-value for your t-test will pop up.
6. However, **you will be required to ALSO report the TEST STATISTIC**
7. **How do you calculate the test statistic?** In a separate cell, go back to the function tab, more functions, statistical and select the T.INV.2T.
8. A new pop up window opens. For the probability, insert the value that you just calculated in the previous steps, and for the df (df=N-2, consult appendix of other handout for further instructions). Then hit enter. **Now you have your t-value, your df and your P-value for your tests**.
    a. You will report (t=3.91, df=18, p=0.001). Note here we will report the p-value being equal (rather than an inequality, as this is the true P-value for this distribution). If your P-value is larger than 0.05, you will still employ the inequality (e.g. t=0.04, df=18, p>0.05).
9. *NOTE: We could have used a critical value table like we did in the correlation analysis. However, we wanted you to have the opportunity to first learn how these tables are used (as described above) and why these values are important. Additionally, we wanted to give you the opportunity to use software to obtain precise P-values which is what you will be doing in upper division biology courses.*