Guide to Excel

Authors: K. N. Carter, B. D. Lamp, D. L. McCurdy and J. M. McCormick*
Last Update: December 26, 2011 by S. Warner and V. Pultz

Introduction

This web page is designed to give the average student a brief introduction to Excel and provide a quick reference to those students who are more familiar with the program. It supplements, but does not replace, hands-on experience. **This tutorial was written for Excel 2010, but other versions are similar, although where you find functions on the menu bar are different.** If you encounter trouble, use the online help, or ask a more experienced user.

Cell References

All cells in a spreadsheet are referred to by their column (letter) and row (number). For example, cell A1 is the first cell in the upper left-hand corner of the worksheet.
Many times it is easier to enter a formula once and then copy and paste it into other cells. When we do this, Excel will automatically change the row and column references. If we don't want the reference to a cell to change, we need to use an *absolute reference*. An absolute reference is denoted with the "$"symbol, and may proceed the column or row designation, or both. For example, if we wanted to subtract the value in cell B13 from that in cell A2, but keeping the reference to cell B13 constant, we might enter in cell C2 the formula "=A2-B13". If we then copy cell C2 and paste it into cell C3, cell C3 would contain the formula "=A3-B13". But if we pasted it into cell D2, cell D2 would contain the formula "=B2-B13".

Basic Mathematical Operations

Formulas may be entered with any combination of numbers and cell references using the standard symbols shown in Table 1. Note that all formulas begin with the equals sign. So, entering "=A1+B1" in cell C1 would add the contents of the two cells and place that result in C1, while entering "=A1+5" in C1 adds 5 to the value in A1 and places it in C1.
The order of operation is important! Multiplication and division are performed before addition and subtraction. Use parentheses to ensure proper calculation. Convince yourself that the order of operations is important by entering "= (2+3)/4" in one cell and "=2+3/4" in another. These calculations do not give the same result!

=	Equals (starts all formulas entered in Excel)
+	Addition
-	Subtraction
/	Division
*	Multiplication
^	Power

Table 1. Symbols for basic mathematical operations in Excel

Built in Functions

Pushing the "Insert Function" button in the "Formula" tab will give access to all of the built-in functions. These functions all require an argument (i. e., the *arg* in the parentheses), which can be either a list of numbers, cell references separated by commas, or two cell references separated by a colon to denote a range of cells. Some useful functions are shown in Table 2, and examples of how they are used are given in Table 3.

AVERAGE(*arg*)	Calculates the average of a set of numbers.
COS(*arg*)	Calculates the cosine of the argument (in radians).
COUNT(*arg*)	Counts the number of cells that contain numbers in the specified range.
COUNTA(arg)	Counts the number of cells that are not empty in the specified range.
EXP(*arg*)	Raises e to the power of the argument.
LN(*arg*)	Returns natural logarithm (ln) of argument.
LOG(*arg*)	Returns the logarithm of argument; defaults to base 10 if no base is specified
LOG10(*arg*)	Returns base 10 logarithm (log) of argument.
PI()	Returns the number pi (this function has no argument).
SQRT(*arg*)	Takes the square root of the argument.
STDEV(*arg*)	Calculates the standard deviation of a set of numbers.
SIN(*arg*)	Calculates the sine of the argument (in radians).
SUM(*arg*)	Adds all the numbers in a range of cells.
TINV(α, *N*-1)	Calculates the value of *t* at a given confidence limit and degree of freedom (*N*-1). Note that (1-α) is the fractional confidence limit (e. g., for the 95% confidence limit, $\alpha = 0.05$).

Table 2. Some useful built-in Excel functions.

Starting with Excel 2010 Microsoft introduced T.INV.2T to replace TINV. The new function may provide improved accuracy. The TINV function is available in Excel 2010 for backward compatibility, but it may not be available in future versions of Excel.

=AVERAGE (A2:A5)	Calculates the average of cells A2, A3, A4 and A5.
=AVERAGE (A2,A4,A5)	Calculates the average of cells A2, A4 and A5.
=SUM(A2:A5)	Adds all the numbers in cells A2, A3, A4, and A5.
=EXP(-1*A2)	Raises e to the -A2 power (note that writing -A2 instead of -1*A2 may not work in older versions of Excel).
=TINV(0.05,COUNT(A2:A10)-1)	Calculates the value of t at the 95% confidence limit for the data contained in cells A2 to A10, inclusive. Note the use of the COUNT function to determine the degrees of freedom.

Table 3. Examples of some Excel functions as they would appear in a spreadsheet cell.

Selecting Cells

To select a single cell, simply click (note that unless otherwise noted "click" means to momentarily depress the left mouse button) on it. To select more than one cell in the same row or column, click on the first cell, and while still holding the left mouse button down, drag the mouse to the last cell. The selected cells will be highlighted. (You can also select a block of cells by clicking on a corner cell and dragging the mouse to the diagonal corner.) An alternate way to select cells is to click on the first cell, hold the *Shift* key down and then select the last cell containing data. Once the cells are selected; let up the *Shift* key. If you want to select an entire column or row, click on the column or row header. To select multiple columns or row, simply click on the header and drag. This last method of selecting cells is very useful if you have a large amount of data that needs to be formatted.

Non-adjacent cells can be selected by holding the *Ctrl* key while clicking on each cell. Highlighting non-adjacent rows or columns is a combination of the steps for selecting non-adjacent cells and rows/columns. Once the first column (or row) has been highlighted, the second column is selected by pressing and holding the *Ctrl* key while you click with the left mouse button in the first cell. While holding the *Ctrl* key and the left mouse button, drag to select the values in this column. The other way to select the second column is to press and hold the *Ctrl* key while clicking in the first cell, then let up the *Ctrl* key and press and hold the *Shift* key. Click on the last cell in the column that contains data and all the cells in between will be highlighted.

Formatting and Toolbar Functions

Many formatting functions are depicted as icons in the "Home" tab along the top of the spreadsheet. The formatting functions are grouped in a category (such as "Font" or "Alignment"). If the function you want is not displayed, there is an expansion button in the bottom right corner of the category which, when clicked, gives more formatting functions. For the "Font," "Alignment," and "Number" categories, this expansion button brings opens the "Format Cells" window. (see Fig. 1: Format Cells Dialog Box) For the "Clipboard" category, it opens the clipboard. If your Excel window is not maximized to full-screen, the function categories become buttons instead of displaying any of the functions. To access the functions,

click the category's button. Some of the more commonly encountered formatting and toolbar functions are given in Table 4. You will note that some of these tools are the same as in word processing programs and that some have menu bar buttons or shortcut keys (e. g., to make something bold you can either click the bold icon in the use the key sequence *Ctrl-b).* To perform any one of the various formatting options, select the object or cell that you wish to format, then use the appropriate shortcut or click the desired formatting button in the "Home" tab. Several formatting functions are also available in the menu displayed when a cell is right-clicked. It is often more efficient to select multiple objects or cells to format before performing the actual formatting. To access the right-click menu with multiple cells selected, select the desired cells then right click one of the selected cells.

Button	Function	Shortcut	Menu Bar Location
≣	Align left	*Ctrl-l*	*Home, Alignment*
≣	Align center	*Ctrl-e*	*Home, Alignment*
≣	Align right	*Ctrl-r*	*Home, Alignment*
B	Bold	*Ctrl-b*	*Home, Font*
I	Italic	*Ctrl-i*	*Home, Font*
U	Underline	*Ctrl-u*	*Home, Font*
.00	Increase number of decimal places		*Home, Number*
.00	Decrease number of decimal places		*Home, Number*
↶ ▾	Undo last action(s)	*Ctrl-z*	*Window Title, Undo*
↷ ▾	Redo last action(s)	*Ctrl-y*	*Window Title, Redo*
A↓Z	Sort cells in ascending order		*Data, Sort & Filter*
Z↓A	Sort cells in descending order		*Data, Sort & Filter*

Table 4. Common tool bar function buttons.

Clicking on any object and selecting the appropriate *format* from the respective tab and category allows you to change the formatting of the object. Right clicking on a selected object and choosing the format option will also bring up the format dialog box. To get the superscript (or a subscript), select the cell you wish to edit. The contents of the cell will appear in the formula bar. Use the mouse to highlight the text to be superscripted (or subscripted), right-click, then select "*F*ormat Cells…" and click on the superscript (or subscript) check box. Another way to do this same formatting is to highlight the desired text and click the *Format Font* dialog box launcher(⊡ See Fig. 1) and click the superscript (or subscript) check box.
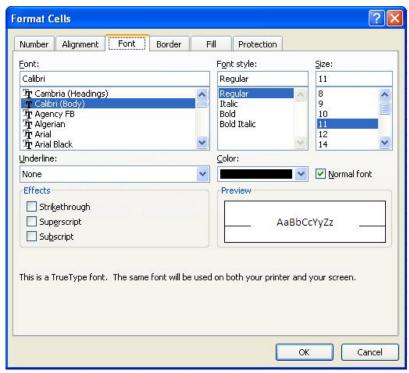
Figure 1: "Format Cells" dialog Box (font tab selected)

To put symbols in axis labels or titles, use the same procedure, except change the font to Symbol font. The correspondence between the Greek letters (Symbol font) and the Roman alphabet (Normal font) for selected symbols is given in Table 5. Symbols that do not appear in Table 5 (such as ±, •, ≠, ≤, ≥, Å, Â, Ä, Ã, Ñ, …, etc.) can be added by selecting "Symbol" (Ω) from the *Insert* tab. Clicking on the appropriate symbol and then clicking the Insert button will complete the task. When finished inserting symbols, close the Symbol window.

Symbol	Roman	Symbol	Roman	Symbol	Roman	Symbol	Roman
α	a	ν	n	A	A	N	N
β	b	o	o	B	B	O	O
χ	c	π	p	X	C	Π	P
δ	d	θ	q	Δ	D	Θ	Q
ε	e	ρ	r	E	E	P	R
ϕ	f	σ	s	Φ	F	Σ	S
γ	g	τ	t	Γ	G	T	T
η	h	υ	u	H	H	Y	U
ι	i	ϖ	v	I	I	ς	V
φ	j	ω	w	ϑ	J	Ω	W
κ	k	ξ	x	K	K	Ξ	X
λ	l	ψ	y	Λ	L	Ψ	Y
μ	m	ζ	z	M	M	Z	Z

Table 5. The relationship between symbol and Roman fonts.

The number of significant figures displayed can be changed using the increase/decrease decimal place button ($^{+.0}_{.00}$ / $^{.00}_{+.0}$) in the *Home* tab and *Number* category. One can also select the Format Font dialog box launcher (⌐ See Fig. 1) and *Number* tab to increase/decrease the number of decimal places displayed. The latter method also allows you to change all aspects of how the numbers are displayed. Note that Excel uses the format "0.00E+00" for scientific notation, not "0.00 x 10+00" or "0.00 x 10^{00}". If you enter a number in the latter formats, Excel will treat it as text.

Sorting

To sort data you must first select the cells to sort. Go to the *Data* tab and click on the Sort button (⌷) in the *Sort & Filter* category. If you have selected one column and there are data in adjacent cells Excel will give a warning message. If you <u>C</u>ontinue with the current selection, the selected cells will be sorted and the adjacent cells will <u>not</u> be sorted. Do <u>not</u> select this option if your data are (*x*, *y*) data pairs, because this will ruin the pairs. If you want to sort data pairs, select the cells containing both values and then sort. When several rows or columns have been selected, using the toolbar buttons such as $^{A}_{Z}$↓ causes the left-most column or top-most row to be sorted. If you use ⌷, you will be prompted for which row/column to sort by.

Graphing

To graph data in the spreadsheet, first select the data. Excel assumes that the *x*-values are in the first column/row selected and that the *y*-values are in the second and any subsequent column/row selected. Now select the *Insert* tab. In the *Charts* category, you will be able to select which type of chart is most appropriate. For most cases in chemistry, select the "Scatter"(⌷) drop-down menu and "Scatter with only Markers" as the graph type.[1] The Chart will be generated automatically and you will then be able to use the *Chart Tools* tabs. See http://chemlab.truman.edu/DataAnalysis/PreparingGraphs_files/PreparingGraphs.asp for more information on graphing.

The first *Chart Tools* tab is *Design* where you can revise your chart type, manipulate the data selected, and modify the layout of your chart. The *Layout* tab is where you can insert photos/shapes/text, add chart labels (chart title, axis titles, and legend labels), and turn gridlines on or off. The *Format* tab is where you can change colors of data selections and chart components and change the size of your chart. Your charts should be on separate excel sheets from your data. To do this, select your chart and go to the *Design* tab and *Location* category. There you will find a "Move Chart" option. With your chart selected, click on "Move Chart" (⌷); a "Move Chart" dialog box will appear. Select *New Sheet* for the chart location. Once a graph has been created it can be edited by simply right-clicking the chart object (axis, data series, etc), and choosing *Format* from the drop-down menu.

A trendline can be added to a graph by selecting the *Layout* tab and in the *Analysis* category click the *Trendline* button. In chemistry we usually want a Linear Trendline. If there is more

than one data set on the graph, you can choose which data set requires a trendline. If you click on <u>M</u>ore Trendline Options… you can check the boxes labeled "Display Equation on chart" and "Display R-squared value on chart". If you have a trendline but do not have its equation displayed, right-click the trendline and click "Format Trendline". From the "Format Trendline" dialog box, select the "Display Equation on chart" check box. Generally we will <u>not</u> force the line to go through a particular value on the y-axis, so leave the "Set Intercept" box unchecked. Once you close the dialog box, Excel will perform a linear least squares fit of the selected data. Note that while a trendline will give you the equation of the best-fit line through the data, it is usually insufficient for the data analysis that is often required in chemistry. Please see the Regression Package section below for a more powerful means of fitting data.

It is possible to copy a chart from Excel and paste it into another program. Make the chart that you wish to copy be the active chart, and while holding the shift key, click on *Copy* () in the *Home* tab and *Clipboard* category. When in the other program, paste the object by clicking the *Paste* button in the *Home* tab of Word, right clicking and selecting a *Paste* option, or use keyboard shortcut *Ctrl*+V. Once in Word, the image size can be modified to fit the page and further edited by selecting the *Format Picture* tab on the ribbon.

Regression Package
In addition to the trendline, Excel also has a more complete regression package, which gives much more information than the simple trendline treatment. This package is part of the *Analysis ToolPak* add-in that can be activated in Excel. Depending on your computer, you may need to install the *Analysis ToolPak* before using the regression tools described here. The installation varies depending on the version of Excel that you are using. Version-specific instructions are below. NOTE: The <u>*Data Analysis*</u> option will only appear if the active window is a worksheet; it will <u>not</u> appear in the menu bar if the active window is a chart.
- If you are using **Excel 2010 or Excel 2007**, select the *Data* tab. If *Data Analysis* appears as an option in the *Analysis* category, the *Analysis ToolPak* is already installed and the regression package can be accessed by clicking on *Data Analysis* and in the pop-up window selecting *Regression* and clicking OK. If *Data Analysis* is not an option on the *Data* tab, follow the instructions for [Excel 2010](#)[2] or for [Excel 2007](#)[3] to install the *Analysis Toolpak*.
- If you are using **Excel 2003**, select the <u>*Tools*</u> menu. If <u>*Data Analysis*</u> appears as an option, the *Analysis ToolPak* is already installed. To use the regression package, select <u>*Tools*</u>, <u>*Data Analysis*</u> from the menu bar and in the pop-up window select *Regression* and click OK. If these statistical tools are not present, follow the instructions [here](#)[4] to install the *Analysis Toolpak*.
- **Mac Users:** The most recent version of Excel for the Mac does not have the Analysis Toolpak built in. Microsoft suggests the following alternative, which functions nearly identically to the Analysis Toolpak. You can use StatPlus:mac LE to perform many of the functions that were previously available in the Analysis ToolPak, such as regressions, histograms, analysis of variance (ANOVA), and t-tests. First, visit the [AnalystSoft](#)[5] web site, and then follow the instructions on the download page. After you have downloaded and installed StatPlus:mac LE, open the workbook that contains the data

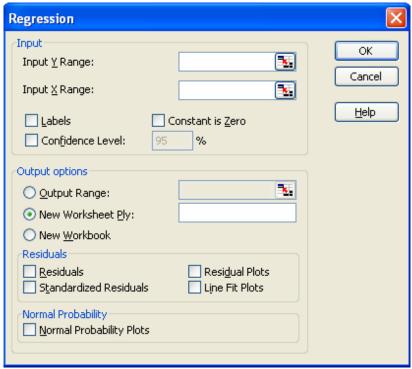that you want to analyze. Then open StatPlus:mac LE. The functions are located on the StatPlus:mac LE menus.

Figure 2. Pop-up window for configuring the regression routine in Excel.

When the regression package is selected, the window shown in Fig. 2 will appear. Input the _Y range_ and _X range_ (if you push the button in the regression window, the window will minimize and you can choose the desired range by clicking and dragging). Because we do not generally want to force a fit to go through zero, leave the check box for _Constant is Zero_ unselected. By clicking on the check box next to _Confidence Level_ and entering a new number in the input box, one can select a new confidence limit. For our purposes this is best left at 95%. Leave the "_Output Options_" set for a _New Worksheet Ply_. Press _OK_, and Excel will perform the regression and output the results to a new worksheet, an example of which is shown as Fig. 3.

The new worksheet contains a good deal of statistical information, but for a basic analysis we need be concerned with only a few items. The first is the value of R^2 (_R Square_), which is a measurement of how related the two measured quantities are ($R^2 = 1$ is a perfect linear relationship). For many of the fits that you will encounter in chemistry $R^2 > 0.999$ are common, and so the use of this parameter to determine how good the fit is or to compare two fits, is somewhat limited. In this case, the standard deviation about the regression, which is given as the _Standard Error_ under the _Regression Statistics_ heading, is more useful.

The values of the intercept and slope (_X Variable 1_) are under the heading of _Coefficients_, while the estimated standard deviation of each parameter is given under the heading _Standard Error_.

The *Lower 95%* and *Upper 95%* entries give the confidence interval for a parameter at 95% confidence. If you selected a confidence level other than 95% the last two columns (the second set of columns in Fig. 3 labeled *Lower 95.0%* and *Upper 95.0%*) will contain the confidence interval at the level. You can calculate the uncertainty (at the *xx%* confidence level) by subtracting the entry under *Lower xx%* from the *Upper xx%* and dividing by 2. For the example given in Fig. 3, the uncertainty in the slope is (2.727447482 - 1.662430005)/2 which is rounded to one significant figure, giving ±0.5 as the answer.

Important! If you are performing a <u>propagation of error analysis</u>[6] on a value calculated from a regression analysis, you do <u>not</u> use the uncertainty! Instead, you must use the standard error (i. e., the standard deviation), <u>not</u> the confidence interval. For example, if we used the slope shown in Fig. 3 in a calculation and we wanted to see how its uncertainty affected the calculated value, we would use ±0.2 (the standard error in the slope rounded to one significant figure) <u>not</u> ±0.5!

If an intercept of zero was specified in the regression window, or if an error occurred, the error "#N/A" will appear for some entries.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.999509156							
R Square	0.999018554							
Adjusted R Square	0.998992726							
Standard Error	0.005716936							
Observations	40							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1.264204688	1.264204688	38680.37433	9.01275E-59			
Residual	38	0.001241968	3.26834E-05					
Total	39	1.265446655						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.329484975	0.002183057	-150.9282695	2.08155E-54	-0.333904342	-0.325065608	-0.333904342	-0.325065608
X Variable 1	-0.00049891	2.53675E-06	-196.6732679	9.01275E-59	-0.000504045	-0.000493775	-0.000504045	-0.000493775

Figure 3. Typical output of the Excel regression package.

Excel's regression package also has a number of <u>advanced features</u>[7] that you may read about, if you wish.

Printing
The appearance of hardcopy can be viewed by going to the *File* tab and selecting *Print*. On the right hand of the screen now is a preview of what your printed page would look like. To adjust page settings, such as margins, go to the *Page Layout* tab. You will need to experiment with combinations of page orientation (portrait or landscape), scaling, margins, etc., until you have the page that matches what you want. If possible, a table should fit completely on a single page. When satisfied with the page settings, return to the *Print* menu under the *File* tab, and push the large "Print" button.

Miscellaneous Features

Excel has a "fill" feature that is useful when one wants to construct a set of data where the *x* values are all equally spaced between some starting and ending value. For example, if we wanted a graph of the function e^{-x}, we could create the *x* values for the graph using the fill function. To fill a series first type the starting value in the first cell of the series and select it by clicking on that cell. Next, in the *Home* tab in the *Editing* category, select the *Fill* button ( ). Select whether you want the created series to occupy <u>R</u>ows or <u>C</u>olumns under the "*Series in*" heading. Under the heading "*Type*", select <u>L</u>inear, which is the most commonly encountered way to fill a data series. Finally, select the <u>St</u>ep value and the <u>Sto</u>p value and hit *Okay*. Excel will fill in the column or row starting at the value initially by adding the step value to it until the stop value is reached.

Error bars, which show the uncertainty in a given point, may be added to a graph in Excel, by first going to the *Layout* tab, *Analysis* category, and selecting *Error Bars*. There are three default options. For more options, click "More Error Bar Options…" You will then be able to select among various options for how the error bars appear (on both sides, one side, or none) and how large to make them. The <u>F</u>ixed value option is good when all of the data have approximately the same uncertainty, while <u>P</u>ercentage assigns the uncertainty as a certain percent of the *y* values. This author has never found the <u>S</u>tandard deviation option to give any useful depiction of the uncertainty, but it may be useful in certain situations. The *Standard error* assigns the uncertainty based on the standard error, which seems to be very similar to the fixed value option. The final option of having <u>C</u>ustom error bars is very useful when each point has a different uncertainty. You can always simply enter whatever value you want in the boxes, but by pushing the button you can specify a range of cells in your spreadsheet that contains your uncertainties. For example, if you have a set of volumes in cells B3 through B10 and their estimated uncertainties from a propagation of error analysis in cells C3 through C10, you would only need to enter the cell references for the uncertainties in the boxes, and each point has its own error bar.

References

1. Dodd, J. S., Ed. *The ACS Style Guide: a Manual for Authors and Editors*; The American Chemical Society: Washington, DC, 1986.

2. Instructions for installing the *Analysis Toolpak for Excel 2010*. http://office.microsoft.com/en-us/excel-help/load-the-analysis-toolpak-HP010342659.aspx?CTT=1 (last visited January 10, 2012).

3. Instructions for installing the *Analysis Toolpak for Excel 2007*. http://office.microsoft.com/en-us/excel/HP100215691033.aspx (last visited January 10, 2012).

4. Instructions for loading the *Analysis Toolpack* in Excel.
http://office.microsoft.com/en-us/excel-help/load-the-analysis-toolpak-HP001127724.aspx?CTT=1
(last visited January 10, 2012).

5. AnalystSoft web site http://www.analystsoft.com/en/products/statplusmacle/ (last
visited January 10, 2012).

6. Propagation of Error on ChemLab.Truman.
http://chemlab.truman.edu/DataAnalysis/Propagation%20of%20Error/PropagationofError.asp
(last visited January 10, 2012).

7. Advanced Regression with Microsoft Excel.
http://chemlab.truman.edu/DataAnalysis/Excel_Files/AdvancedRegression.asp (last
visited January 10, 2012).