Spreadsheet Help
ver. 27 July 2023
by
Brian M. Tissue

Abstract

This document provides an introduction to the use of a spreadsheet for analyzing and plotting scientific data. The screen captures are from Microsoft Excel 2007™, but newer software versions operate in the same way as described here. Examples and step-by-step instructions are provided for common mathematical operations, statistical tests, data plots, and linear regression. Use this guide to help you develop your own spreadsheets. The "You-Try-It" guides also provide templates for many common chemistry calculations. For general background on statistics I suggest the text by Taylor.[1] For advanced Excel techniques and examples I use the book by de Levie.[2]

The most current version of this document is available online at:
> **http://www.achem.org**.

This website and associated files are provided by the author to assist users of the text:
> B. M. Tissue, *Basics of Analytical Chemistry and Chemical Equilibria,* 2nd Ed. (John Wiley: New York, 2023).

Contents   **Page**

---

[1] John R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd Ed., University Science Books: Sausalito, CA (1997).

[2] Robert de Levie, *Advanced Excel for scientific data analysis*, 2nd ed., Oxford University Press: New York (2008).

I.Introduction

Spreadsheets are sufficiently powerful to perform most data analysis and plotting needs in analytical chemistry. Advantages of a spreadsheet include:
- rapid calculations of large data sets or repetitive operations
- instant update of results after editing data or formulas
- graphical plotting of data sets and results
- transfer of data and plots to other electronic documents

The disadvantage is that you need a computer and you must learn the "language" of Excel to write formulas. Simple calculations will always be quicker to do with a hand calculator or in your head. Calculations in a spreadsheet can be less error-prone and more efficient than doing them on paper, but validating that a calculation is correct is also more difficult when working with formulas. It is prudent to work calculations on paper first to validate that you get a sensible answer and that units cancel, especially for multi-step operations.

The specific instructions in this document apply to Microsoft Excel 2007™. Functions and conventions for other programs or operating systems might be different. I use the "⇒" arrow to indicate making a choice from a ribbon or menu. For example, Insert ⇒ Symbol means to click on the `Insert` ribbon tab and then on the `Symbol` button to show the dialog box to enter special characters.

II.Getting Started

A **spreadsheet** is the electronic file that you save and open with the Excel program. A **worksheet** is the page where you enter data and formulas to perform calculations. At the bottom of the screen are a set of tabs labeled Sheet1, Sheet2, and so on (see Figure 1). Clicking a tab accesses the different worksheets in a spreadsheet. You can rename a worksheet by double-clicking on a tab and typing a new label. Clicking on the rightmost tab will create a new worksheet.
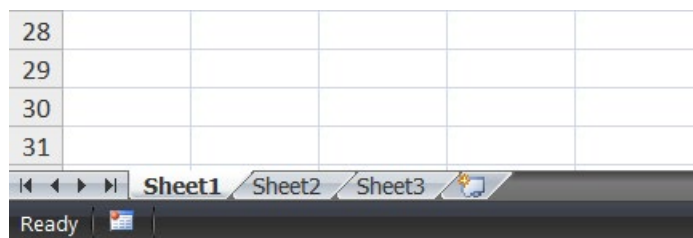

Figure 1. Worksheet tabs.

On starting Excel you will see the Office Button and Quick Access Toolbar at the top of the screen. These buttons access file, print, and other commands. Below the toolbar are tabs to access **ribbons** of command buttons. Figure 2 shows some of the formatting commands on the Home ribbon. Clicking on any of the tabs to the right of the Home tab displays the other ribbons.
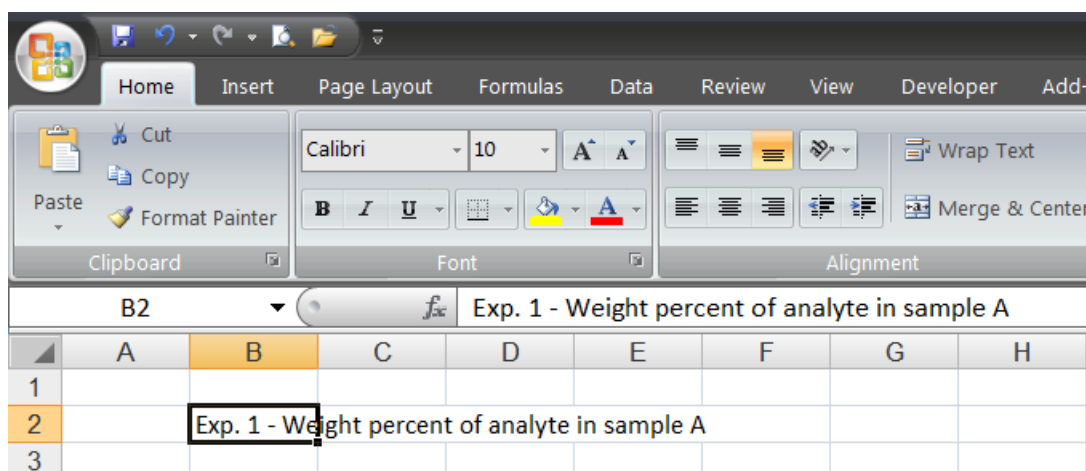
Figure 2. Formatting commands on the Home ribbon.

A new worksheet consists of a grid of empty **cells**. Each cell is specified by its column letter and row number. Data entry and editing is as simple as clicking on a cell and typing. In Figure 2 I clicked on cell B2 and typed a title for the spreadsheet. Note the highlighting of the headings for column B and row 2. What you type will display above the column labels in the **Formula Bar** (visible in Figure 2 next to f_x). You can move the cursor within the formula bar with the keyboard arrows or by inserting it with the mouse to edit text, numbers, and formulas. Figure 3 shows the result after I have clicked on cells B4 through C8 to add text and data values.

**Figure 3. Data entry in a worksheet.**

Text and data can also be copied into cells from another electronic file. Highlight the text to be copied and press the Ctrl-C keyboard combination to copy the data to the clipboard. Return to an Excel worksheet, click on a cell, and press Ctrl-V to paste the copied data. Copy and paste will not place multiple columns of text data into separate worksheet columns.[3] If your source data is in multiple columns, open the file using the Text Import Wizard. We'll step through an example of this process in Section VI.

---

[3] Copying data from a table in Word will usually paste correctly, but copying from multiple columns in text or pdf files will not.

Note in Figure 3 that the text in cell B4 is obscured by the text in cell C4. To adjust the column widths to make everything visible, hover the mouse over the dividing line between column headings. When the cursor changes to a vertical line, click to "grab" the column and move the mouse to change the width. Widths of multiple columns can be adjusted by selecting multiple column headings with the mouse and then adjusting one column to the desired width. Widths are also adjustable by selecting one or more columns, right-clicking on a column heading, selecting `Column Width...' and entering a specific width in the dialog box. Figure 4 shows the result. I've also made the text boldface using the formatting buttons on the Home ribbon.

An important issue is illustrated in cells B6 and C8. I typed 1.1460 and 0.5380 in these cells, but the display does not show the trailing zero. Spreadsheets are like calculators in that they do not keep track of significant figures. Format the cells containing numbers to show them with four decimal places, which is the precision in this experimental data. Do this by selecting the cells with the mouse and right-clicking (see Figure 4). In the pop-up menu select `Format Cells...' and a dialog box will pop up (see Figure 5). In the dialog box, select `Number' from the left menu and then enter the desired number of decimal places.

WORK ALONG: Open Excel and enter the text and values as shown in Figure 3. The data may be typed or copied from the appendix at the end of this document. Save the file.
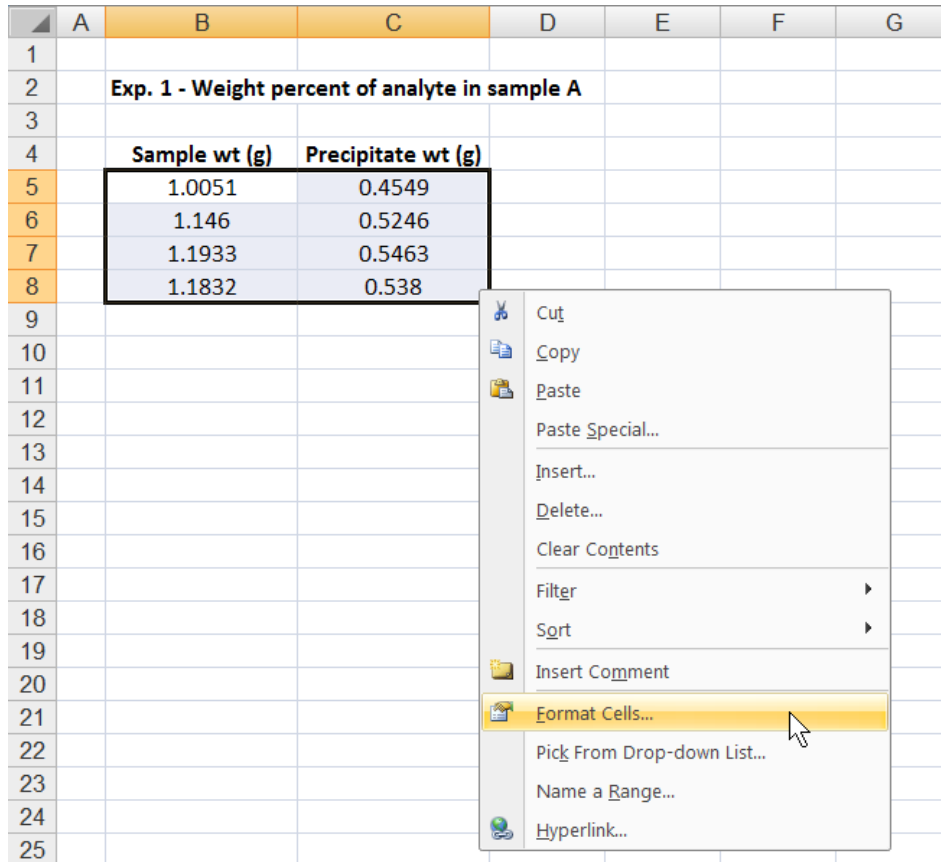
Figure 4. Selecting cells and right-clicking to access menu.
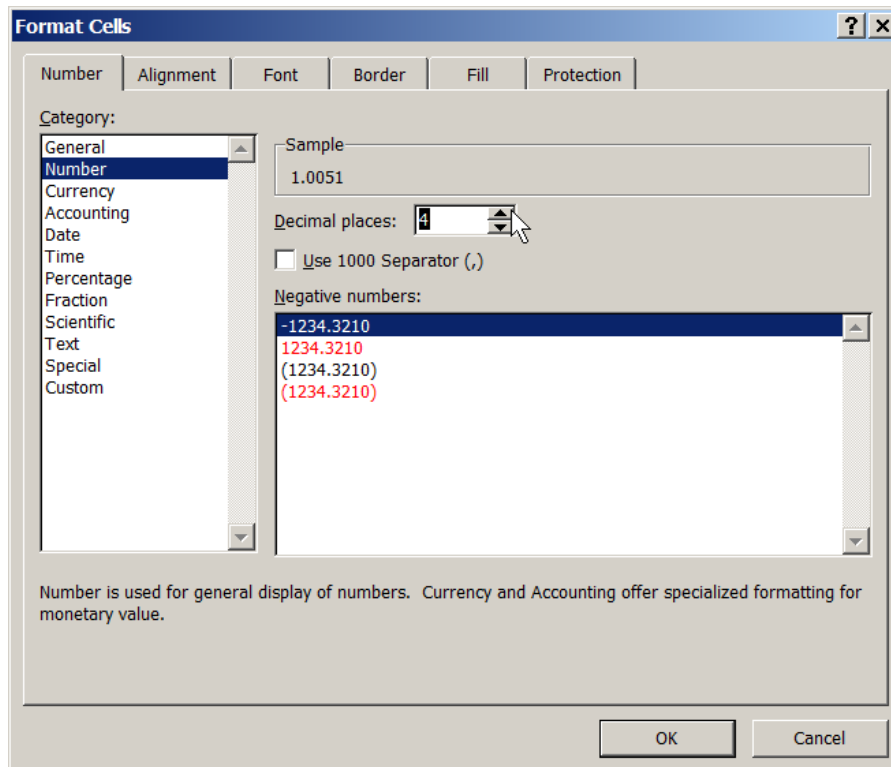
Figure 5. Cell formatting dialog box.

Table 1 lists the most common formats for mathematical calculations. Numbers that vary by orders of magnitude can be displayed using scientific notation. If you click on a cell and type 5e2, the cell will display 5.00E+02, meaning $5×10^2$. Use `Format Cells…' to specify the number of decimal places in `Scientific' notation.

Table 1. Numerical Cell Formats

General:	Excel autoformats text and numbers for display
Number:	number with a set number of digits past the decimal point
Percentage:	multiplies cell value by 100 and displays % symbol
Scientific:	exponential format with set number of digits past the decimal point
Text:	displays numerical values as entered

The default cell format of General (see **Figure 5**) will usually be sufficient to display cell contents as text or numbers. Formatting cells can be tedious, so it is not necessary to format cells for intermediate steps in a calculation. It is important to format the number of decimal places to specify the precision in your original data and in the final results.

Now with data entered and formatted, I will type a formula in cell E5 to calculate the weight percent for the first measurement. **Formulas** in Excel start with the "=" sign followed by the mathematical operation. Referring to a particular cell is done with the letter+number combination.[4] For the data in row 5, I type in cell E5:
    =(C5/B5)*100
This formula takes the number in cell C5, divides it by the number in cell B5, and multiplies the result by 100.

After entering this formula cell E5 shows the numerical result. Copy it by clicking on cell E5 and pressing the Ctrl-C keyboard combination to copy it to the clipboard. Now I highlight cells E6, E7, and E8, and press Ctrl-V to paste the formula into these cells. The formula automatically changes the cell reference from row 5 to rows 6, 7, and 8. **Figure 6** shows the results. Another way to copy formulas is to highlight the cell containing a formula, grab the corner and drag it down through the cells where you want to copy the formula. Simply highlighting a cell and double-clicking on the lower right corner with the mouse pointer also copies a formula down a column when it calls values from another column of data, as this example does.

---

[4] Excel also supports a relative cell referencing format. To change click on the Office Button ⇒ Excel Options ⇒ Formulas ⇒ and check or uncheck the "R1C1 reference style" box. The formula examples in this document assume that this setting is unchecked.

◢	A	B	C	D	E	F
1						
2		Exp. 1 - Weight percent of analyte in sample A				
3						
4		Sample wt (g)	Precipitate wt (g)		Wt-%	
5		1.0051	0.4549		45.2592	%
6		1.1460	0.5246		45.7766	%
7		1.1933	0.5463		45.7806	%
8		1.1832	0.5380		45.4699	%
9						
10					45.5716	%
11					0.2541	%
12					0.5575	%-RSD

Figure 6. Formulas to find the average, standard deviation, and %-RSD of the data set.

Now type formulas in cells E10, E11, and E12 to determine the average, standard deviation, and relative standard deviation (the formulas are shown in **Figure 7**). The colon in the notation In these formulas specifies performing an operation using the data in a **range of cells**. In this case `E5:E8' is equivalent to typing the list `E5, E6, E7, E8'.

It is a good practice to include some documentation on your worksheet. The amount of explanatory text will depend on the use of the spreadsheet. If you will be sharing it with coworkers, you will need enough information for someone else to decipher and edit the results. At a minimum, you will want enough information so that you know the source of the data and you can understand the formulas if you reuse the spreadsheet in the future.

In my example, I had already typed a title, so I also add my name, date, and labels. Since Excel does not maintain units, I type the unit symbols next to the numbers in column F. If you are sharing a worksheet with someone else, it is also useful to show the formulas that are used to obtain the results. Simply typing a formula attempts to calculate a numerical result. To force a formula to display as text, precede the formula with an apostrophe, e.g., in cell G10 type:
    '=AVERAGE(E5:E8)

Error! Reference source not found. shows the final worksheet. I've added borders, shading, bold-face, and centered some of the cell contents to improve the overall organization and readability. As with documentation, the amount of time that you will spend on formatting will depend on the use of the spreadsheet. If you will copy and paste parts of a worksheet into a formal report, you will want to take a more critical look at improving the readability so that other users will be able to interpret your results.

	A	B	C	D	E	F	G	H
1								
2		Exp. 1 - Weight percent of analyte in sample A					B.M. Tissue, 05 Jul 2010	
3								
4		Sample wt (g)	Precipitate wt (g)		Wt-%		Formulas	
5		1.0051	0.4549		45.2592	%	=(C5/B5)*100	
6		1.1460	0.5246		45.7766	%		
7		1.1933	0.5463		45.7806	%		
8		1.1832	0.5380		45.4699	%		
9								
10				mean:	45.5716	%	=AVERAGE(E5:E8)	
11				std dev:	0.2541	%	=STDEV(E5:E8)	
12			relative standard deviation:		0.5575	%-RSD	=(E11/E10)*100	
13								
14					ANSWER: 45.57 ± 0.25 %			
15								

Figure 7. Formatting and documentation in a worksheet.

> ***WORK ALONG:*** Add the formulas for mean, standard deviation, and %-RSD to your worksheet. Check your values with Figure 7. Add labels and formatting to highlight the results.

More Functions and Formulas

Excel has an extensive set of functions to perform many types of mathematical operation. If you know the name of the operation that you want, you can simply type the function into a formula. If you don't know the Excel name for an operation, click on the **Formulas Ribbon** and look through the menus. Figure 8 shows a screen capture of part of the Formulas ribbon. The most useful functions for our calculations are under the Math & Trig and the Statistical menus. If you do not find an operation or you need help on how to use a function, click on the question mark at the upper right of the Excel screen to search the **Excel Help** for information.
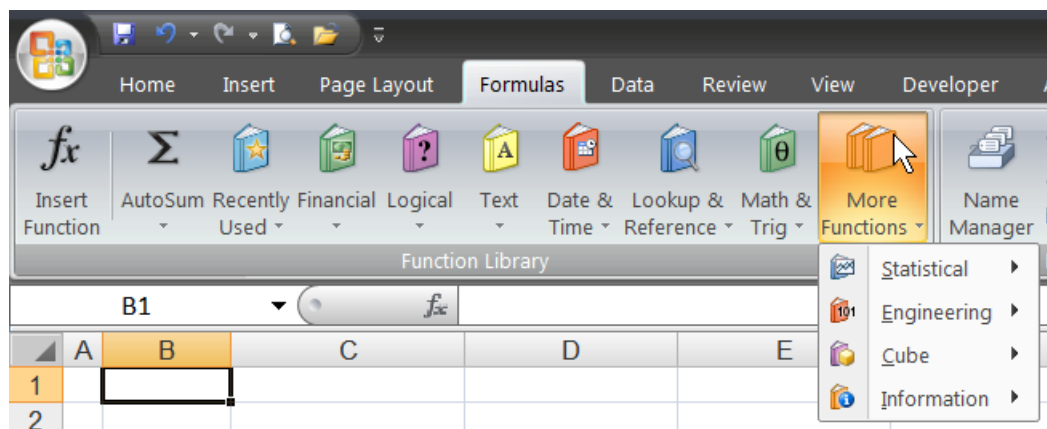

Figure 8. The Function Library on the Formulas ribbon.

Table 2 lists a small sampling of built-in math functions for common operations. The "number" in a formula can be a typed value, a cell reference, or a sub-formula. A full list of functions is available in the Excel Help.

Table 2. Useful Excel commands for mathematical formulas

function	Excel command	
addition (+)	=number+number	
subtraction (-)	=number-number	
multiplication (*)	=number*number	
division (/)	=number/number	
raise to a power (^)	=number^(number)	
raise number to any power	=POWER(number,power)	(same as =number^power)
base 10 logarithm	=LOG(number)	
natural logarithm	=LN(number)	
raise e to a number	=EXP(number)	
take square root	=SQRT(number)	
the value of pi	=PI()	
find absolute value	=ABS(number)	
find sine of an angle	=SIN(number)	(other trig functions also built-in)

You can use mathematical operators in any combination in a formula, but it is often necessary to include some operations in parentheses to specify the order of operation. For example, Excel does division before addition so the formulas =55+33/22 and =(55+33)/22 give different results. For some of the common mathematical operations, the order of precedence is:
1. negation (−)
2. exponentiation (∧)
3. division (/) and multiplication (*)
4. addition (+) and subtraction (−)

If there is any doubt about the order of operations in a formula, the safe approach is to enclose parts of the formula in parentheses. Note that many of the functions consist of the function name and one or more parameters enclosed in parentheses. In these cases the parentheses are required. Table 3 lists several examples.

Table 3. Examples of operator precedence.

w/o parentheses	with parentheses	w/o parentheses	with parentheses	w/o parentheses	with parentheses
=55+33/22	=(55+33)/22	=-3^2	=-(3^2)	=exp(-4-3/7)	=exp(-(4-3)/7)
55+1.5 = 56.5	88/22 = 4	$-3^2 = 9$	$-(3^2) = -9$	$e^{-4.43} = 0.012$	$e^{-1/7} = 0.867$

WORK ALONG: In a new worksheet enter the Table 3 formulas with and without parentheses and compare your values to the listed results.
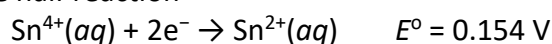
As a detailed example, we'll write the formula for the Nernst equation, which is used to predict half reaction voltages when reactants and products are not at standard conditions:

$$E = E^o - \frac{RT}{nF} \ln(Q)$$

where:

- E^o is the standard reduction potential
- R is the molar gas constant, 8.3145 J mol^{-1} K^{-1}
- T is absolute temperature
- n is the number of electrons transferred in the half reaction
- F is the Faraday constant, 96485 C mol^{-1}
- Q is the reaction quotient for the half reaction

For the half reaction
$$Sn^{4+}(aq) + 2e^- \rightarrow Sn^{2+}(aq) \qquad E^o = 0.154 \text{ V}$$
The Nernst equation becomes:

$$E = E^o - \frac{RT}{nF} \ln \frac{[Sn^{2+}]}{[Sn^{4+}]}$$

I've used concentration rather than activity symbols in this expression for clarity. The expression in Excel format is:

=E^o-(R*T/(n*F))*ln(top/bottom)

Rather than typing values in formulas, it is usually better to use a cell location to be able to change a value easily. Figure 9 shows the locations where I've placed the constants. I will perform the calculation as a function of relative concentration of [Sn^{2+}] and [Sn^{4+}], which are placed in columns E and F. On replacing the variables with cell locations, the formula in cell G5 now becomes:

=$C$6-$C$7*$C$8/($C$9*$C$10)*LN($E5/$F5)

	A	B	C	D	E	F	G
1							
2		**Nernst Equation calculation**					
3							
4		Sn^{4+} + 2e$^-$ → Sn^{2+}			[Sn^{2+}]	[Sn^{4+}]	E (V)
5					0.0001	1	0.273
6		E^o =	0.154	V	0.001	1	0.243
7		R =	8.314	J/mol/K	0.01	1	0.213
8		T =	299.0	K	0.1	1	0.184
9		n =	2		1	1	0.154
10		F =	96485	C/mol	1	0.1	0.124
11					1	0.01	0.095
12					1	0.001	0.065
13					1	0.0001	0.035

Figure 9. Nernst equation calculations.

Omitting the parentheses before the LN function gives an erroneous result:
=$C$6-$C$7*$C$8/$C$9*$C$10*LN($E5/$F5)

Writing the formula in the following way does produce a correct result:
=$C$6-$C$7*$C$8/$C$9/$C$10*LN($E5/$F5)

WORK ALONG: In a new worksheet type the constants for the Nernst equation and replicate the table of results in Figure 9. Rewrite the formula with different sets of parentheses.

III. Statistical Functions

Section II illustrated the use of AVERAGE(range) and STDEV(range). Table 4 lists these and other Excel functions that are useful for statistical calculations. In addition to these functions, there are a number of more specialized functions that evaluate text, e.g., COUNTA, or that include a criterion to process cells, e.g., COUNTIF. Information on using these types of functions can be found in the Excel Help database.[5]

Table 4. Useful Excel functions for statistical analysis.

operation	Excel function	
number of data points	=COUNT(range)	
summation of a data set	=SUM(range)	
mean of a data set	=AVERAGE(range)	
sample standard deviation	=STDEV(range)	
minimum value in a data set	=MIN(range)	see also SMALL
maximum value in a data set	=MAX(range)	see also LARGE

Excel also has Descriptive Statistics that are accessed from:
> Data ⇒ Data Analysis

If you do not see Data Analysis, load the Analysis Toolpak as an Add-In.

To load the Analysis Toolpak:
1. Click the Microsoft Office Button ⇒ Excel Options ⇒ Add-Ins (left menu)
2. In the `Manage:' box near the bottom, select `Excel Add-ins' and click the `Go...' button.
3. Check the box next to `Analysis ToolPak' and then click the OK button.
4. The Data Analysis command will now be in the Analysis group on the Data ribbon.

The main advantage of Descriptive Statistics is that you don't have to worry about making a mistake in a formula. A disadvantage is that the values returned by Descriptive Statistics are not updated if you change the data. On changing data you must run the Descriptive Statistics again. Because of the lack of updating, I recommend that you write your own formulas to obtain statistical results.

Clicking `Data Analysis' brings up a list of tools. Click on `Descriptive Statistics' and then the OK button. A dialog box will appear where you enter the cell range of your data in `Input Range:'. Under Output options check `Output Range:' and enter a cell to place the results. Check the box for `Summary statistics' and click the OK button.

---

[5] Click on the small round question mark button at the upper right of the Excel screen.
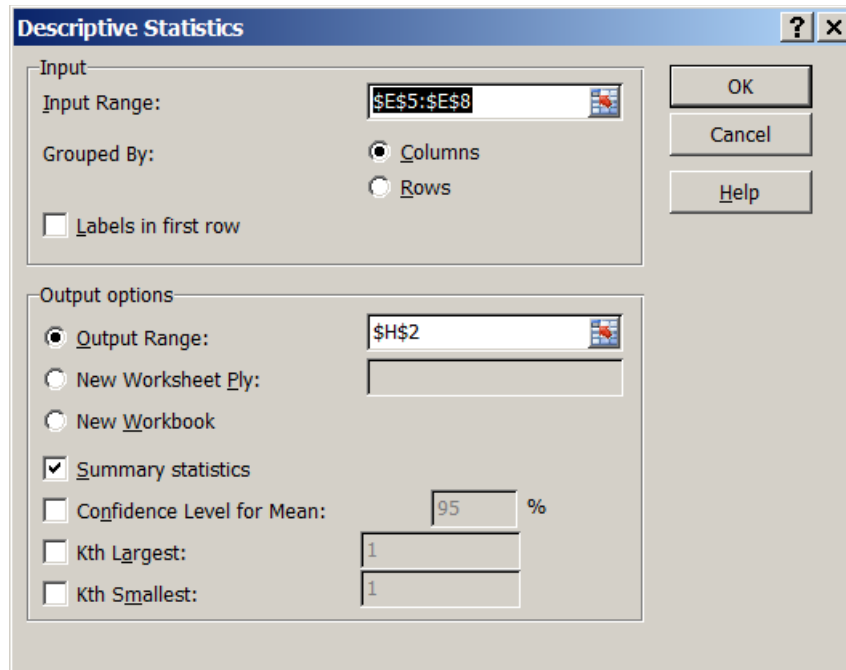
Figure 10. Descriptive Statistic dialog box.

Figure 11 shows an example of the results that the Descriptive Statistics returns. The Mean is at the top and the Standard Deviation is in the fifth row down. These tools work fine, but Excel provides no explanation of the returned values.

| A	B	C	D	E	F	G	H	I
1								
2	**Exp. 1 - Weight percent of analyte in sample A**						*Column1*	
3								
4	**Sample wt (g)**	**Precipitate wt (g)**		**Wt-%**			Mean	45.57157825
5	1.0051	0.4549		45.259	%		Standard Error	0.127037787
6	1.1460	0.5246		45.777	%		Median	45.62326321
7	1.1933	0.5463		45.781	%		Mode	#N/A
8	1.1832	0.5380		45.470	%		Standard Deviation	0.254075573
9							Sample Variance	0.064554397
10							Kurtosis	-2.85525641
11							Skewness	-0.560355798
12							Range	0.521430206
13							Minimum	45.25917819
14							Maximum	45.7806084
15							Sum	182.286313
16							Count	4

Figure 11. Descriptive Statistics results.

IV. Plotting Data

Before continuing to methods to analyze x,y data sets, it is useful to discuss some basic tips on plotting data. Figure 12 shows [Ca^{2+}] measurements of river water samples using an ion-selective electrode. Two samples were taken from each location along the river. The average and standard deviation of the duplicate measurements are found in columns E and F, respectively.

|	A	B	C	D	E	F	G
1							
2		Nosuch River Water Analysis					
3		Table 1. Calcium Potentiometry Measurements					
4			Sample A	Sample B	average		
5		Location	ISE (mV)	ISE (mV)	ISE (mV)	Std Dev (mV)	dashed line
6		NR-8902	39.9	39.5	39.7	0.28	39.00
7		NR-8903	36.3	36.6	36.5	0.21	39.00
8		NR-8904	39.8	40.1	40.0	0.21	39.00
9		NR-8905	39.0	39.4	39.2	0.28	39.00
10		NR-8906	39.6	39.8	39.7	0.14	39.00
11		NR-8907	39.2	38.8	39.0	0.28	39.00
12							
13		Measurement Notes:			Data Summary		
14		procedure taken from ISE manual			N =	6	
15		100 mL water samples			sum =	234.0	
16		2 mL ionic strength adjuster buffer			mean =	39.00	
17		Ca2+ ion-selective electrode			std dev =	1.30	
18		Ag/AgCl reference electrode					
19		all solutions at room temp (23.5 C)					
20		See B.M. Tissue lab notebook II, pg 23-24.					

Figure 12. ISE measurements of river water.

To plot a series of measurements of one variable, select the data as shown in Figure 12. Click on the `Insert' ribbon and choose the type of chart that you want. For repetitive data, a Column, Line, or Bar chart is appropriate. Figure 13 shows a partial view of the drop-down menu for creating a Column chart.
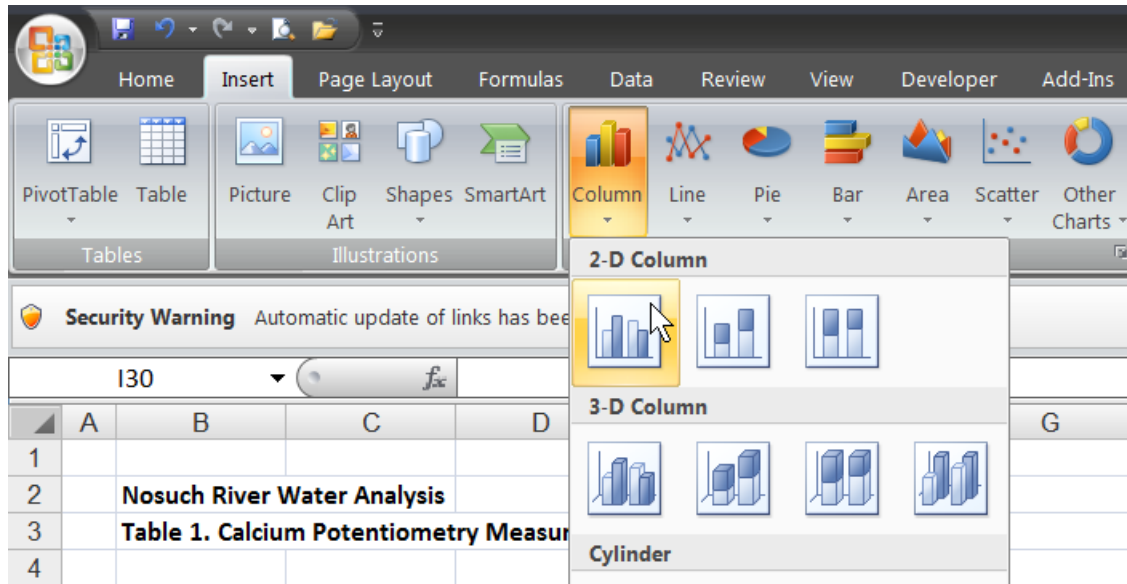
Figure 13. Inserting a chart for graphical display of data.

After clicking on the 2-D Column button, the default chart will look something like Figure 14. We will make a number of changes to improve the look of this chart. The series label is unnecessary for a single data set and can be deleted. To adjust how a chart displays, right click on the part that you want to change and choose `Format...' from the pop-up menu.
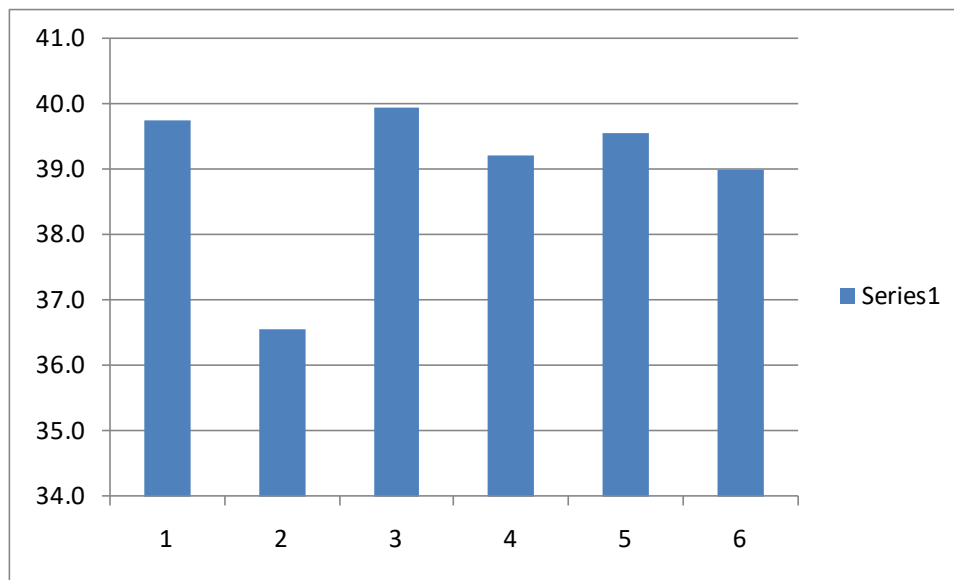

Figure 14. A 2-D Column chart.

If you cannot access a title or heading, click on the chart so that the `Chart Tools' menus appear with the other ribbon tabs. The `Layout' ribbon has buttons to add `Axis Titles' and other chart options (see Figure 15). The `Design' and `Format' ribbons provide other commands to change the appearance of a chart.
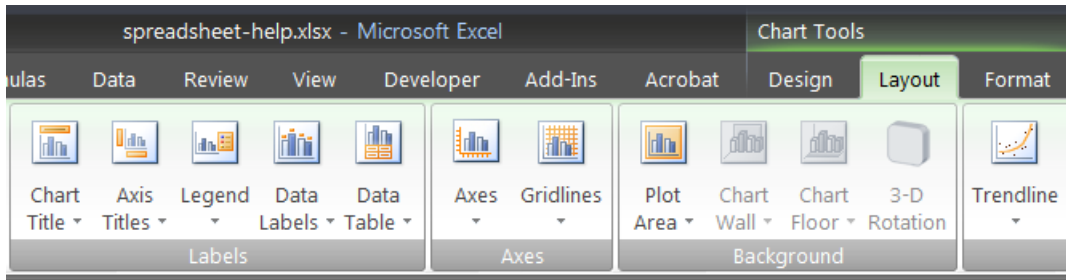
Figure 15. A portion of the Layout ribbon.

Excel autoscales the y axis by default. The scatter in a series of data measurements can be misleading depending on the selected scale. Right-click on an axis and select `Format Axis…` to change the minimum and maximum values. Figure 16 shows the data with the y axis minimum set to zero, additional axis titles, and other format changes. The scatter in the data is less obvious in this plot.
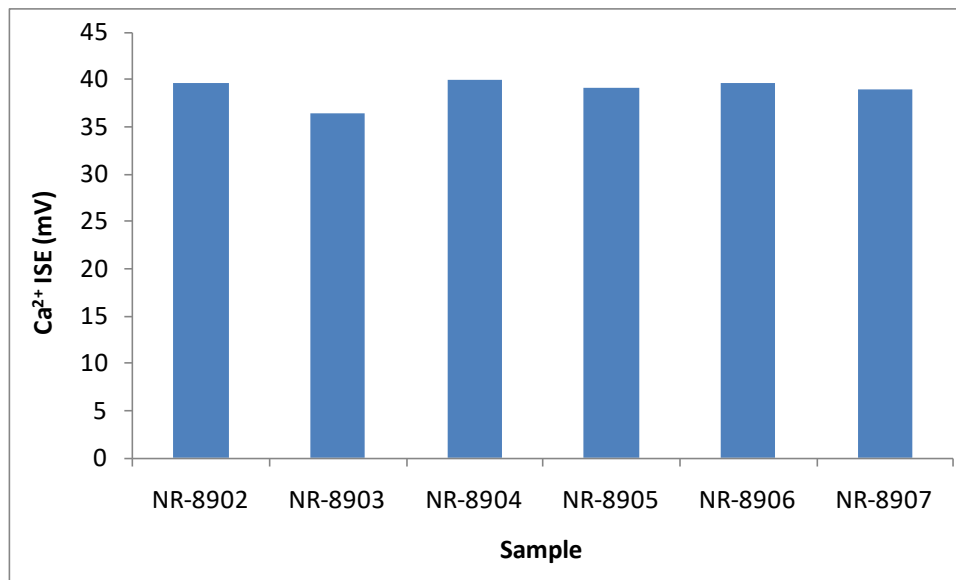
Figure 16. Chart with full scale y axis.

Figure 17 shows an alternate format where I've added a chart title, data labels, and other text. The error bars are taken from column F. The dashed line is made from the average, which is copied to column G (see Figure 12).
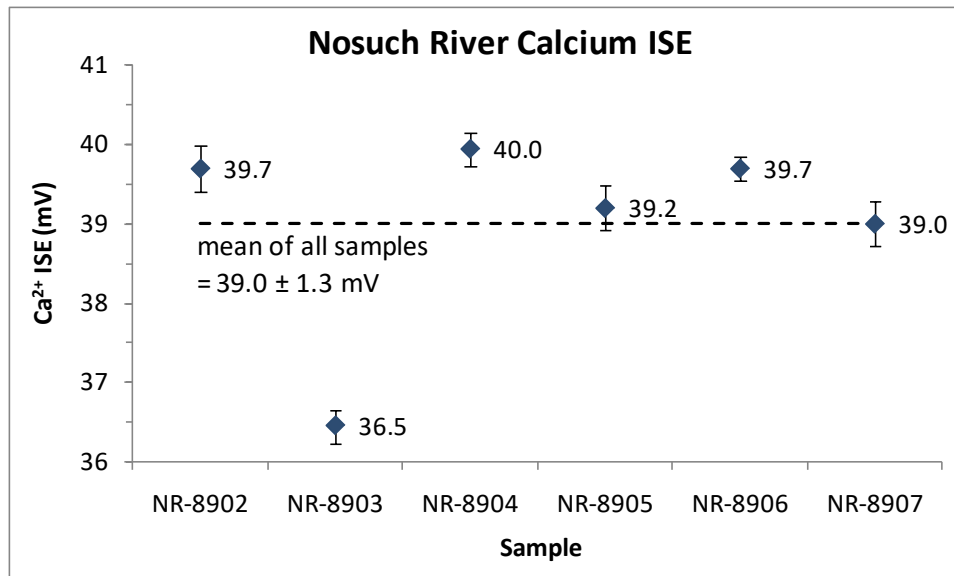
Figure 17. Another way to format a chart.

> **WORK ALONG:** In a new worksheet type, or paste from the appendix, the data from Figure 13. Select the data and insert a 2-D Line chart. Right click on the line in the chart and choose `Format Data Series...`. Change various parameters for Marker Options, Line Style, etc. Add axis titles and adjust other chart options in the Design and Format ribbons to see the effects.

V.Linear Regression

Many analytical methods are calibrated by measuring an instrument signal versus the concentration of a series of known standards. This process provides a calibration or working curve to determine the analyte concentration in unknown samples. The equation that describes the calibration data is called the **calibration function**. This calibration function is found by determining the parameters to an equation that is the "best fit" to the calibration data. This process is called **least-squares fitting** or **linear regression** for the case of fitting the parameters a_0 and a_1 to the linear equation:

$$y = a_0 + a_1 x$$

Excel provides a number of methods to perform linear regression on x,y data. Table 5 lists four of these methods with selected characteristics.

Table 5. Excel methods of linear regression.

Method	Accessed by	Cell Reference	Updates	Standard Deviations	Functions
trendline	chart	no	yes	no	multiple
SLOPE/INTERCEPT	function	yes	yes	no	linear
LINEST	function	yes	yes	yes	linear
Regression	Data Analysis	yes	no	yes	linear

To compare these methods briefly, trendline is quick and simple and can fit data to other functions besides a linear equation. It displays the best-fit calibration function on a chart, so the parameters must be retyped to use in calculations. The SLOPE and INTERCEPT functions provide the a_0 and a_1 parameters in cells that can be referenced for other calculations. LINEST and Regression provide more information, including the precision of the parameters. Regression is called from the Data Analysis tools, and must be run again if the data changes. When using any of the methods besides trendline, it is advisable to plot the data and residuals to check that the data points are linear over the entire range of x values.

Sample linear data is available in the appendix and online at:

> http://www.itl.nist.gov/div898/strd/lls/data/Norris.shtml[6]

On this web page, right-click on the link `Data file (ASCII Format)' and use `Save Target As…' to save the file to your computer. In Excel, click the Microsoft Office Button ⇒ Open, and navigate to the location where you saved the file. If you do not see Norris.dat, check that the file filter is set for `All Files(*.*)'.

Click the Open button and the Text Import Wizard will appear. Choose `Delimited' for file type and click the Next button. At step 2 choose `Space' as a delimiter and click the Finish button. You can delete much of the imported text. Note that the data is listed with y values first. Copy

---

[6] Norris, J., *Calibration of Ozone Monitors*, National Institute of Standards and Technology, Gaithersburg, MD.

the y values to the right of the x values before doing a plot. Use this data to perform linear regression using Trendline and the SLOPE and INTERCEPT functions.

Trendline

To plot a set of x,y data pairs, select the two columns of data and use:
>   Insert ⇒ Scatter (in the Charts group).

Right click on a data point in the chart and choose `Add Trendline…' from the pop-up menu. The `Format Trendline' dialog box will appear as shown in Figure 18. Choose `Linear' for the `Trend/Regression Type' and check the boxes for `Display equation on chart' and `Display R-squared value on chart'. Click on the Close button and a line and equations will now display with the data. The result is shown in Figure 19.

To adjust the format of the equations on the chart, right-click on an equation and select `Format Trendline Label…'. Adjust the number of decimal places and click the Close button.
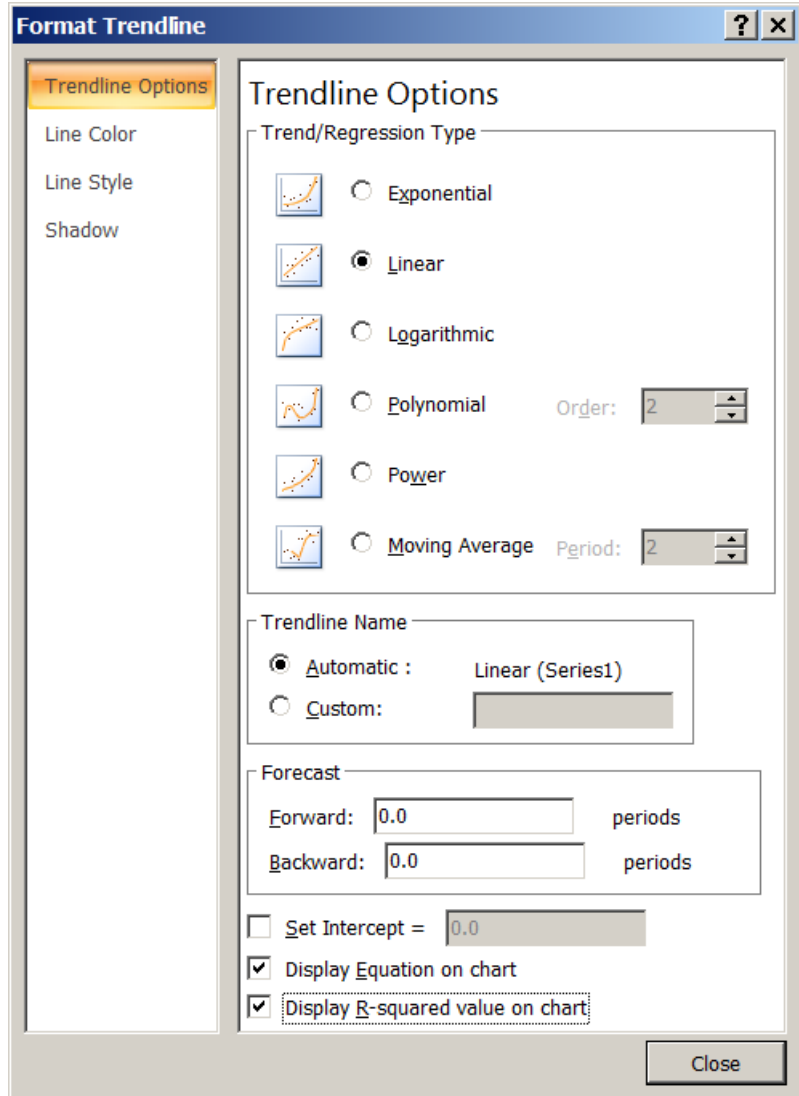
Figure 18. Trendline dialog box.

Trendline is the quickest way to get the linear equation for a set of x,y data. The disadvantage of Trendline is that we must retype the slope and intercept parameters in cells to use them in calculations. If we change the x,y data, the Trendline will update the equation on the chart but not the values typed in cells. The next section shows how to use the SLOPE and INTERCEPT functions to place the best-fit parameters in cells to use in calculations.
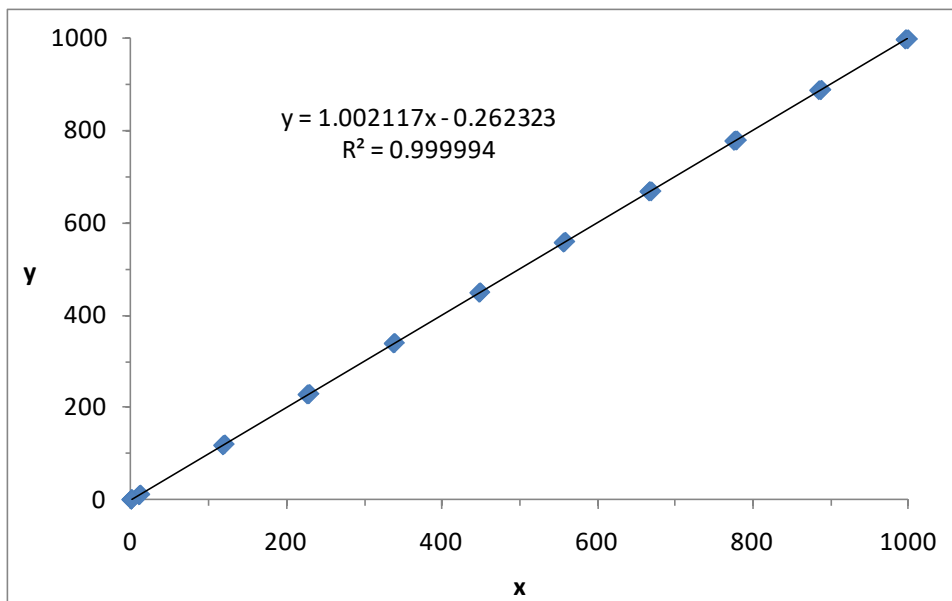
Figure 19. A set of x,y data with trendline.

> **WORK ALONG:** In a new worksheet import the Norris x,y data. Insert a scatter chart and add a trendline.

Using SLOPE and INTERCEPT to calculate deviations

A deviation or residual, d_i, is the difference between an experimental y value and the value calculated from the calibration function using the best-fit parameters. Plotting the deviations is useful to ensure that the chosen calibration function matches the experimental data. A plot of deviations versus x should show a random distribution around zero. Any curvature or pattern in a plot of deviations indicates that the calibration function does not match the experimental data. The deviation for each x value, d_i, is calculated as:

d_i = (experimental y value) − (calculated y value)

or

$d_i = y_i - (a_0 + a_1 x_i)$

We can use INTERCEPT(y-values, x-values) and SLOPE(y-values, x-values) to place a_0 and a_1 in cells to calculate d_i. Figure 20 shows the case of x and y data values in columns B and C, respectively. In cells E4 and E5 I've typed:

    =SLOPE($C$7:$C$43,$B$7:$B$43)
    =INTERCEPT($C$7:$C$43,$B$7:$B$43)

where the arguments in parentheses are the ranges of the y and x data, respectively. These formulas should return the same values that appear in the Trendline equation on the chart. In cell D8 and E8 I've typed the following formulas to find the calculated y value and the deviation:

    =B8*$E$4+$E$5
    =C8-D8

I select cells D8:E8 and double-click the selection handle (the small square on the lower left corner of the border) to copy the formulas down the columns. You can also select cells D8:E8, press Ctrl-C, and paste the formulas in the cells below. The top few rows of results are shown in Figure 20.

	A	B	C	D	E	F
1						
2		**Linear Regression of Norris data.***				
3						
4				slope =	1.002117	
5				intercept =	-0.262323	
6						
7		*x*	*y*	*y* calc	*d* i	
8		0.2	0.1	-0.06	0.16	
9		337.4	338.8	337.85	0.95	
10		118.2	118.1	118.19	-0.09	

Figure 20. Results using the SLOPE and INTERCEPT functions.

Figure 21 shows the deviations for the Norris data plotted as a Scatter chart. In this case there is no trend to suggest that the data is not linear. To create this chart I selected the x data in column B. I then held down the Ctrl key and select the deviation values in column E. Once the columns of data are selected, a Scatter chart is chosen from the Insert ribbon.
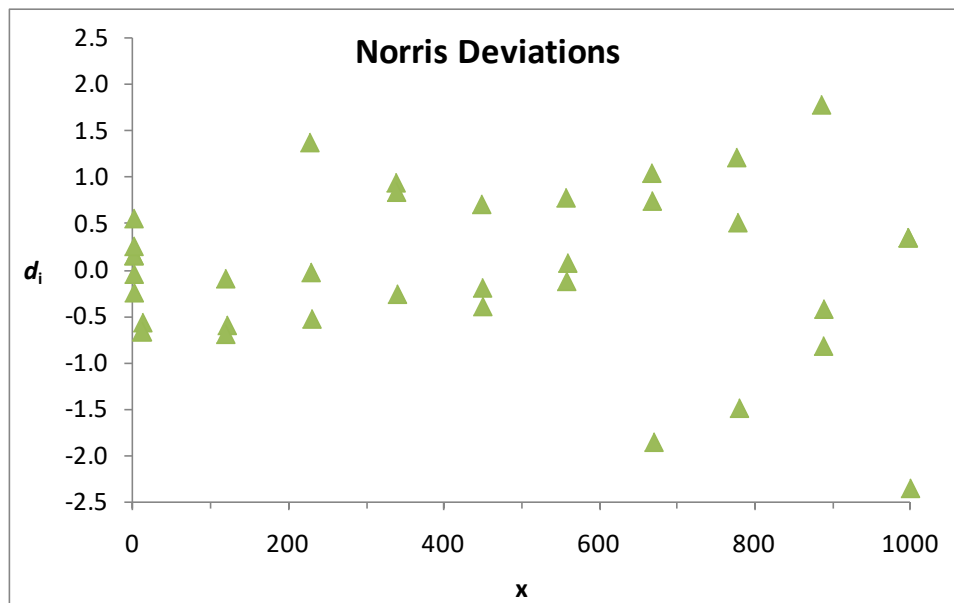

**Figure 21. Deviations for the Norris data.**

Deviations can also be highlighted, copied, and pasted directly on the original chart. When doing so, it is usually necessary to right click on a residual data point, select `Format Data

Series...', and choose `Secondary Axis'. The second axis on the right side of the chart can now be scaled and formatted independently from the primary y axis.

> **WORK ALONG:** Use the SLOPE and INTERCEPT functions on the Norris data and compare the results to the Trendline values. Write formulas in new columns to generate calculated y values and residuals. Plot the calculated y values and the residuals in a single chart.

VI.Summary

This document has provided several examples of common operations for analyzing analytical data. There are two primary sources to obtain more help on Excel functions or to customize the Excel program. Clicking on the question mark at the upper right of the Excel screen brings up the **Excel Help** dialog box (see Figure 22). You can browse by help category or search for specific topics.
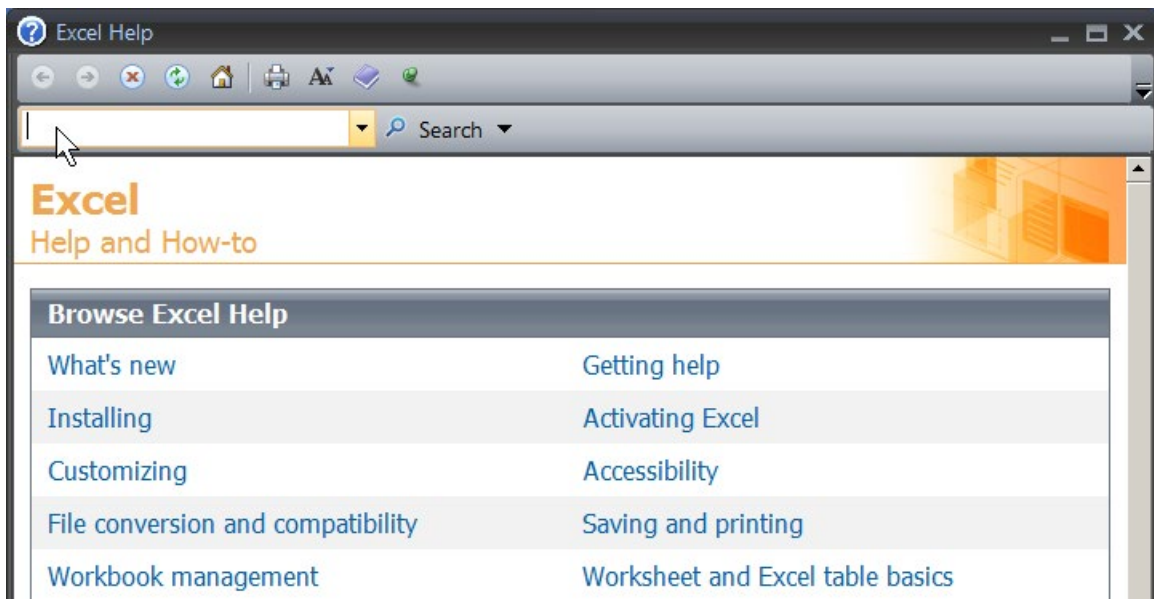
Figure 22. Excel Help.

Another source of information is the Microsoft Office website. You can search for information at https://support.microsoft.com/en-us.

VII. Appendix

The first two data sets are in separate columns so they can be easily copied into Excel. To maintain the two columns of the third data set, copy the data to a text file, save the file, and open the file from Excel. The data can also be downloaded from the URL in the footnote.

Data for Section II. Getting Started

```
Sample wt (g)
1.0051
1.1460
1.1933
1.1832
```

```
Precipitate wt (g)
0.4549
0.5246
0.5463
0.5380
```

Data for Section V. Plotting Data

```
Sample A (mV)
39.9
36.3
39.8
39.0
39.6
39.2
```

```
Sample B (mV)
39.5
36.6
40.1
39.4
39.8
38.8
```

Data for Section VI. Linear Regression

```
NIST/ITL StRD  Dataset Name:  Norris (Norris.dat)
Procedure:     Linear Least Squares Regression
Reference:     Norris, J., NIST, Calibration of Ozone Monitors.[7]

                  Certified Regression Statistics

                                      Standard Deviation
      Parameter         Estimate          of Estimate
         a0       -0.262323073774029    0.232818234301152
         a1        1.00211681802045     0.429796848199937E-03
```

y	x
0.1	0.2
338.8	337.4
118.1	118.2
888.0	884.6
9.2	10.1
228.1	226.5
668.5	666.3
998.5	996.3
449.1	448.6
778.9	777.0
559.2	558.2
0.3	0.4
0.1	0.6
778.1	775.5
668.8	666.9
339.3	338.0
448.9	447.5
10.8	11.6
557.7	556.0
228.3	228.1
998.0	995.8
888.8	887.6
119.6	120.2
0.3	0.3
0.6	0.3
557.6	556.8
339.3	339.1
888.0	887.2
998.5	999.0
778.9	779.0
10.2	11.1
117.6	118.3
228.9	229.2
668.4	669.1
449.2	448.9
0.2	0.5

---

[7] Norris, J., *Calibration of Ozone Monitors*, National Institute of Standards and Technology, Gaithersburg, MD; http://www.itl.nist.gov/div898/strd/lls/data/Norris.shtml.