LAB 1_____USING EXCEL FOR ANALYZING CHEMISTRY DATA

Introduction

A spreadsheet is a useful tool for organizing raw data and performing basic mathematical and statistical analysis. Learning how to use spreadsheets in the beginning of the semester and continuing to improve your skills will save you precious time performing data analysis after each lab. Although the calculations necessary for the first few labs can easily be done using a scientific calculator, it becomes impractical, if not impossible, to perform data analysis for the labs later in the semester without the use of computerized spreadsheets. Additionally, some of you may think that it is not worth your time to learn how to use Microsoft Excel because you do not plan on pursuing a career in chemistry; however, learning how to compose and manipulate spreadsheets is a useful tool for any type of data organization including social science research, public health research, and medical research.

Some of the major advantages of using a spreadsheet are that repetitive calculations can be duplicated with just a few keystrokes, mistakes in calculations can be fixed by changing single cells or functions, and data is organized in a clear and understandable way.

Tutorial

Rather than simply giving you a list of Excel commands, this 'lab' will walk you through some Excel basics in the context of some simple (and difficult) chemistry problems that you might encounter in a quantitative chemistry lab context. Depending on your level of Excel fluency, you may already know some of the basics. Try to use this tutorial to help develop some of the more elegant and useful commands.

During this tutorial you will be asked to construct a spreadsheet of your own. Make sure to save your work frequently in order to secure your work. Also consider adopting an appropriate naming convention for your lab data files, such as "course_year_last_first_lab#.xlsx" (CH999_2012_doe_j_lab0.xlsx is the name of file that Jane Doe would use for Lab #0 of CH999 in 2012). Answers to the other questions should be given on the post-lab assignment page at the end of the tutorial.

Remember, your lab instructor is available to help you if you get stuck. The important thing is that you finish and get a good grasp of the basics of Excel.

Part 1 - Organization of a spreadsheet

Each box in the spreadsheet is called a *cell* and these cells can be referenced using a combination of a letter and a number. The letters represent the columns of the spreadsheet and the numbers represent the rows of the spreadsheet.

Open up your spreadsheet and locate cell **D5**. This cell is in the fourth column and fifth row of the spreadsheet.

What is the code used to reference the cell in the eleventh column and fortieth row?

Note: All of the commands in this tutorial will be typed in bold to differentiate from the rest of the text, but while performing actual calculations, you should not type the commands in bold font.

Part 2 - Spreadsheet fundamentals

Now that we have a frame of reference for discussing spreadsheets, let's discuss some of the more fundamental aspects/uses of spreadsheets.

One use of a spreadsheet is that it can be used as a basic calculator, but unlike most calculators, it stores your results, which can later be referenced in subsequent calculations. All calculations in the spreadsheet must begin with an equals sign, indicating to the program that you are performing a calculation. Below is a table listing some of the basic mathematical operations and the correlated commands.

Operation	Excel command
Addition	=A1+B1
Subtraction	=A1-B1
Multiplication	=A1*B1
Division	=A1/B1
Exponents	$=$ A1 \land B1

Notice that multiplication must be done using the asterisk (*); Excel does not know how to perform multiplication using "x" or parentheses (i.e. (A1)(B1) would not work). We will now use some of these mathematical operations to solve a basic chemistry question using Excel.

Worked Example #1: If you mix 5.000 grams of $\text{Fe}_2O_3(s)$ with an excess of CO(g), a) how many grams and molecules of Fe(s) will form, and b) how many grams and molecules of $\text{CO}_2(g)$ will form?

$$\operatorname{Fe}_2\operatorname{O}_3(s) + 3\operatorname{CO}(g) \to 2\operatorname{Fe}(s) + 3\operatorname{CO}_2(g)$$

\rightarrow							
Start by generating the following table in your spreadsheet:							
		Α	В	С	D	E	
	1		stoichiometry	grams	moles	molecules	
	2	$Fe_2O_3(s)$	1	5.000			
	3	Fe(s)	2				
	4	CO ₂ (g)	3				
	5	CO(g)	3	excess	excess	excess	
Figure 1.1: Worked Example #1							
Solve for the number of moles of $Fe_2O_3(s)$ (you should get 0.03131 moles of							
Fe ₂ O ₃ (s) in cell D2).							
r in the remainder of the table.							

You may realize that there are an excessive number of decimal places in some of your values. Excel does not know how to account for significant figures by itself, nor do you need to worry about significant figures. Consider Excel to be an overgrown calculator; it isn't relevant how many figures Excel shows. Rather, it is the number that you report in your assignments that is important.

That said, you may want to change the number of decimal places that are being displayed (to make it easier to read or to see more detail). This is accomplished with the *format cells* feature. With this feature it is possible to convert the numbers within the cells to the correct number of significant figures, scientific notation, fractions, percentages, and many other formats that you may need.

Convert values in column D into scientific notation, and all values within the table to have three decimal places (cell **D4** should read 9.393×10^{-2})

With calculators there is always the danger that rounding too early will result in an incorrect answer due to 'roundoff error;' for that reason, you are always advised to not round to the correct number of significant figures until the end of your calculations. That said, one of the big advantages of Excel is that it does not actually round the number that is stored in memory - changing the formatting only limits the precision that it displays.

Now change cell **C2** to 6.000 g. What effect does this change have on the products?

After you have found the results of changing cell C2 to 6.000 g, convert all the values in the table back to correct values when C2 was 5.000 g.

Part 3 - Referencing Cells

When typing out the equations to fill in the blank cells, it is not necessary to retype numbers that are already present, but rather you can click on the reference cell containing that number. Consider cell D4 in Worked Example #1. Your formula should have looked like:

= (5.000/159.69) * (3/1)

where 159.69 is the molar mass of $Fe_2O_3(s)$ and 3/1 is the stoichiometric ratio. Instead of manually entering the numbers, consider the following version that references the appropriate cells in the table:

$=D2^{*}(B3/B2)$

When copying the value from one cell to another, you can click the standard *copy* and then *paste*. If the initial cell does not contain a equation it will simply copy the value. However, if you copy a cell containing an equation, the standard copy and paste sequence copy the formula, and not the value. If you want to copy and paste the value within the cell and not the equation, first click on *copy* followed by *paste special* and then *values*.

If you copy a formula containing reference cells down a number of rows then it will automatically adjust the cell references to match the new cell. For instance, let's assume that you have the formula $=(G2/A2)^*H2$ in cell I2. If you copy this cell to I3 then the formula will now read: $=(G3/A3)^*H3$. If you copy the cell to J3 then the formula will read $=(H3/B3)^*I3$.

It is often the case that we want to copy the general formula, but that some of the cell references should remain unchanged. Adding a dollar sign (\$) before the letter or number in the cell reference then it will remain unchanged when copied. If, in our previous example, cell **I2** had been =(G2/\$A\$2)*\$H2 then copying the cell to **J3** would give the formula: =(H3/\$A\$2)*\$H3.

Make a new sheet within your document by clicking on the plus symbol next to *sheet1*. Rename sheet 1 to WE1a (i.e. worked example 1a) and rename the newly created sheet to WE1b.

On sheet WE1b, redo the table from Worked Example #1 to use referenced cells wherever appropriate. At this point, the output from both tables should be identical.

Change the value in cell **C2** to 6.500. What are the values in cells **E3** and **E4**?

Hopefully you now realize that it is very valuable to reference cells; especially because if you make a mistake in a calculation you can just go back and correct the one cell with the mistake and all subsequent cells that include the reference will adjust according to the correction.

Worked Example #2: You are preparing a set of dilutions from a stock iron (Fe) solution with initial concentration of 0.50 mg/L.

	Α	В	С	D	E
1	Constants	Amount	Unit		
2	[Fe] _{stock}	0.50	mg/L		
3	Temp	25	°C		
4	Temp	298	к		
5	Pressure	1	atm		
6					
	Solution	Volume Stock	Total final	Dilution conc.	
7	number	added (L)	volume (L)	(mg/L)	
8	1	0.0010	0.10		
9	2	0.0050	0.10		
10	3	0.010	0.10		
11	4	0.015	0.10		
Figu	re 1.2: Wo	ked Example $\#$	<i></i> ≢2		
Figu	re 1.2: Wor	eked Example $\#$	±2	D8 is 5.0 ×	10^{-3} W
Figur ete ue c	re 1.2: Worth the table (the table $(t of cell D11?)$	ked Examplg # € he correct va	^{∉2} lue for cell	D8 is 5.0 ×	10^{-3}). W

For the sake of checking your work we will need to display the actual formulas used – so that the teaching assistants for the class can see where the data for the calculations originated. **Note:** we are only doing this for this lab so that we can make sure that the formulas you used are correct; these formulas **do not** count as *sample calculations*¹ for the purposes of grading your reports.

Displaying the equation or formula can be accomplished using the following steps: Start by clicking on the cell until the formula becomes visible. Select the formula and click 'copy.' 'Paste' the formula into the desired cell and insert an apostrophe (') before the equals sign.

Finally, you need to enable the 'Print Row and Column Headings.' In the 'Page Setup' pop-up click 'Sheet' and then select Print 'Row and Column Headings.' Now, when you print it will show the row numbers and column labels. Always make sure to have these headings printed if you submit a spreadsheet.

Fill in column E on sheet "WE2" with the correct formulas.

¹In this course, one full set of sample calculations are required for all problems. These calculations must include units as well as the correct number of significant figures for final answers.

Part 4 - Working some problems

Now you will try to solve some chemistry problems on your own: please solve these problems on paper first, and submit both handwritten and computer versions to your lab instructor. You can use the two methods of solving each problem to check your own work. The answers to problems 1, 2, and 3, should be saved on sheets P1, P2, and P3, respectively. Make sure to display the formulas that you used next to the appropriate columns.

Problem #1: What is the volume of a single water molecule in m^3 ? You may or may not find the following information useful: Density of water is 0.9956 g/mL, Melting Point of water is 32°F, the pKa of water is 15.74, and the molar mass of water is 18.016 g/mol. Proper units have to be used throughout each calculation to receive credit and the final answer must contain the correct units and the correct significant figures.

Problem #2: NaOH(aq) is often used to titrate HCl(aq).

$$\mathrm{HCl}(aq) + \mathrm{OH}^{-}(aq) \rightarrow \mathrm{Cl}^{-}(aq) + \mathrm{H}_{2}\mathrm{O}(l)$$

A 0.500M NaOH solution is used to titrate several aliquots of 25.00 mL of an unknown concentration of HCl. Use the data in the table below and the skills in this tutorial to determine the average concentration of the hydrochloric acid solution and the standard deviation. Provide a handwritten sample calculation (for one trial) in the post-lab assignment.

Trial #	Volume NaOH
	(mL)
1	15.55
2	15.77
3	15.54
4	16.01
5	15.89
6	15.99
7	15.62
8	15.66
9	15.71
10	15.84

Part 5 - Graphing

By now you've hopefully come to appreciate the computing and organizational abilities of Excel. In addition to its excellent ability to keep track of data and perform calculations, Excel can also be used to make graphs. Graphs are an essential tool in the scientist's arsenal - they enable someone to present large amounts of data in a concise manner that immediately draws the reader's attention to a trend or pattern. Unfortunately, the format and types of default graphs that Excel makes are far from on-par for science purposes. Below are a *few* general guidelines for making a good graph:

• Only use "xy scatter" plots (some scientific applications will also use bar graphs or pies)

- Excel can plot the points or it can connect them with lines (smoothed or straight). Rarely will points be connected with lines this implies data that you don't have; the only exceptions are chromatograms and spectra, where you would not include points and only use lines.
- All axes need labels and units (check significant figures of tick mark labels)
- Excel displays vertical axis gridlines (horizontal lines on the whole graph) by default. Either remove all of the gridlines or make a perfect grid on both axes (most graphs should not have gridlines).
- Excel includes a legend on the side of the graph by default. Single data set graphs should not have a legend. If multiple data sets are present (2 or more) then a legend or a caption is needed. If there are only a few sets (2 to 3), use only a caption. If there are many sets (4 or more) then use a legend **and** a caption.
- Unless you use a color printer, do not use color lines to distinguish between data sets on a single graph.
- Resize the axes so that the plotted lines extends across most of the graph (the origin need not be included on the graph)
- Where appropriate (for linear data), add a trendline and display its equation on the graph (include R-squared as well).
- The title of the figure is the first part of the figure caption, not an obnoxiously large title in the top center of the figure (Excel default).

Problem #3: A public health researcher hypothesized that there is a correlation between height and weight. (Recall: 1 kg = 2.2 lbs and 1 cm = 0.39 in)

Weight (lbs)
158
163
153
251
139
201
169

Transcribe the table above to a new sheet (P3). Prepare a graph of Weight (in kg) vs Height (in cm) and move the resulting graph to its own sheet, Graph1. In your post-lab assignment, discuss whether or not the data support the researcher's hypothesis? Explain. (Use three significant figures in this problem)





Figure 1.3: Incorrect version of a graph



4

Figure1. Graph of Absorbance vs Temperature.

Figure 1.4: Correct version of a graph

Reference - Useful Commands

Average

To compute the average value of a set of numbers, use the =AVERAGE(first_cell:last_cell) command.

Standard deviation

Calculating the standard deviation is very similar to calculating the average value of a data set. Just type =**STDEV(first_cell:last_cell)**.

Count cells

If you have a long list of cells and want to determine the N value (number of data points) for future calculations, use the **=COUNT(array of cells)** command. This array can be one or two-dimensional.

Relative standard deviation (RSD)

There is no explicit Excel command for RSD. To calculate the RSD, you should divide the standard deviation by the average using the two values as reference cells.

$t \ statistic$

To compute the student t statistic use the =**TINV(probability, DOF)** command, where DOF is the number of degrees of freedom (N-1) and probability is 1-(confidence limit). Example: for 75 data points at the 95% confidence limit use =**TINV(0.05,74)**.

F statistic

Use the =**FINV**(**probability**, **DOF1**, **DOF2**) to compute the *F* statistic for two data sets, where DOF1 and DOF2 are the degrees of freedom of the numerator data set and denominator data set, respectively, and the probability is 1-(confidence limit). For example, if you have 6 degrees of freedom in the numerator and 4 degrees of freedom in the denominator and want the *F*-value for a 95% confidence limit, then use =**FINV**(.05,6,4).

If Statements

If statements are used to evaluate whether a statement is true or false. If the statement is true the command will return one value, and if the statement is false then the command will return a different value. The command is =IF(logical test, value if true, value if false). For example, if one wanted to determine whether the value in cell C3 is greater than 30, and if it is greater than 30 that means the person is considered moderately obese and if its

less than or equal to 30 the person is only considered overweight, the command would be: =IF(C3>30, "MODERATELY OBESE", "OVERWEIGHT").

Reference - Least Squares Regression

Even though it is convenient to graph on the spreadsheet and be able to automatically receive the linear equation, one deficiency is that the equation displayed in the graph does not contain the appropriate confidence intervals. To calculate the confidence $\sum ntervals$ it is recommended to create a template for a linear equation using the method of least squares.

In Excel, the command for generating the method of least squares values is =LINEST(y-values, x-values,1,1). This command will result in two-by-three (or two-by-five) table containing an output of slope (m), standard deviation of slope (s_m) , R squared, y-intercept (b), standard deviation in y-intercept (s_b) , and the standard deviation in $y(s_y)$. Start by preparing a table exactly like the one in the figure below:

	А	В	С	D
1	m			b
2	Sm			Sb
3	R ²			Sy

Figure 1.5: LINEST Table

Next, highlight the six cells in between the letters and type the **=LINEST** command as described above. DO NOT hit enter! Be sure to type ctrl+enter (PC) or command+enter (MAC) when you finish typing in the command to receive all the desired data. If not, only the first cell will be computed. At this point you now have all of the regression parameters (slope, intercept) and their associated uncertainties. The only value that you are missing that is the uncertainty in value you wish to regress, s_x , which can be solved using the following equation:

$$s_x = \frac{s_y}{|m|} \sqrt{\frac{1}{k} + \frac{1}{N} + \frac{(y - \bar{y})^2}{m^2 \sum_i (x_i - \bar{x})^2}}$$

where k is the number of replicate measurements in x, N is the number of points on the line.

To solve for s_x , your template should also include the command to count the number of x-values (**COUNT**), the mean value for y-values, and the command to calculate $\sum_i (x_i - \bar{x})^2$, which can be calculated using the **=DEVSQ(x-values)** command.

Post-Lab Assignment

Name: _____

Section: _____

Submitting your spreadsheets

In addition to the answers to the following questions, print and submit the properly labeled sheets from your spreadsheet (WE1a, WE1b, WE2, P1, P2, P3, P4, and Graph4) - each sheet should not take more than a single piece of paper, and must have the row and column headings shown.

Make sure that your spreadsheets include units and that the final values are shown to the correct number of significant figures.

Part 1: What is the code used to reference the cell in the eleventh column and fortieth row?

Part 2: Now change cell C2 to 6.0000 g. What effect does this change have on the products?

Part 3:

- Change the value in cell C2 to 6.5000. What are the values in cells E3 and E4?
- Complete the table (the correct value for cell **D8** is 5.0×10^{-3}). What is the value of cell **D11**?
- In your own words, describe what would have happened if you would have only put a dollar sign in front of the B.

Part 4:● Problem #1

• Problem #2

Part 5: Problem #3. Does the data support the researcher's hypothesis? Explain briefly.