Analytical Chemistry
Calibration Curve Handout

I. Quick-and Dirty Excel Tutorial

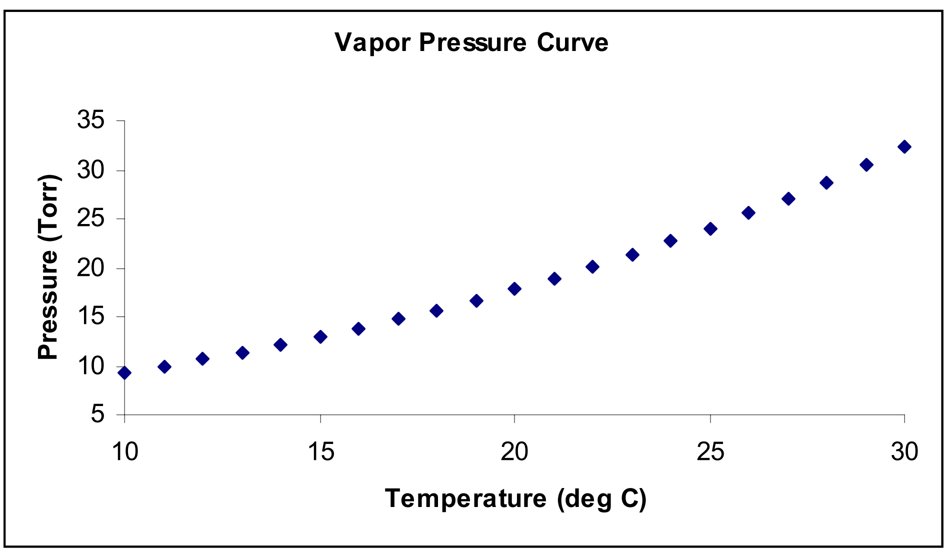
For those of you with little experience with Excel, I've provided some key techniques that should help you use the program both for problem sets and lab write-ups. Please come talk with me if you have questions about any of this material—I've come to love(!) Excel more and more throughout my career, and I'm always happy to help people exploit the program's power.

I also recommend you read Harris' excellent discussions of Excel techniques. In particular, I recommend Sections 2-10 and 2-11 (pp. 38-42), the example on pp. 84-85, and Section 5-5 (pp. 92-93).

Much of the power of Excel comes from the use of cell references. For example, say we wanted to plot the vapor pressure of water in Torr as a function of temperature in ºC. Let's further say I want the plot to go from 10ºC to 30ºC, with points at every degree. (This could be useful in experiments in which we collect gases under water.) We are given the following empirical equation (called the "Antoine equation"):

$$\log P = A - \frac{B}{T + C}$$

In this equation, P is the vapor pressure in atm, T is the temperature in K, and A, B, and C are constants. (In the temperature range of interest, A = 5.40221, B = 1838.675, and C = -31.737.) The vapor pressure therefore depends on a continuously changing variable, T, and on three constants. Excel is great at solving problems like this! The plot is below, and the spreadsheet I used to generate the plot is on the next page.



	A	B	C	D	E	F	G
1	Constants		T (deg C)	T (K)	log P	P (atm)	P (Torr)
2	A =	5.40221	10	283.15	-1.911155	0.01227	9.325212
3	B =	1838.675	11	284.15	-1.882181	0.013117	9.968562
4	C =	-31.737	12	285.15	-1.853436	0.014014	10.65069
5			13	286.15	-1.824917	0.014965	11.37357
6			14	287.15	-1.796621	0.015973	12.13927
7			15	288.15	-1.768546	0.017039	12.94994
8			16	289.15	-1.740689	0.018168	13.80781
9			17	290.15	-1.713047	0.019362	14.7152
10			18	291.15	-1.685619	0.020624	15.67453
11			19	292.15	-1.658401	0.021958	16.6883
12			20	293.15	-1.631392	0.023367	17.75913
13			21	294.15	-1.604588	0.024855	18.8897
14			22	295.15	-1.577988	0.026425	20.08284
15			23	296.15	-1.55159	0.028081	21.34146
16			24	297.15	-1.52539	0.029827	22.66856
17			25	298.15	-1.499386	0.031667	24.06729
18			26	299.15	-1.473578	0.033606	25.54089
19			27	300.15	-1.447961	0.035648	27.09271
20			28	301.15	-1.422535	0.037798	28.72623
21			29	302.15	-1.397297	0.040059	30.44507
22			30	303.15	-1.372244	0.042438	32.25293

Spreadsheet Remarks

Note that I have entered the formula's constants into cells B2, B3, and B4. This will allow me to refer to them when I define formulas, and to re-define them at will. (For example, if I wanted the vapor pressure in a different temperature range, I would need to use a different set of constants.)

In cell C2, I have entered the number 10. Instead of typing in 11, 12, 13, etc. in the cells below, I type the following formula into cell C3: = C2+1 In this formula, C2 is a relative cell reference. What this formula really tells cell C3 is, "Take the number right above you, and add 1 to it."

I can then copy the formula in C3 into cells C4 through C22. Here are two ways to do this:

(1) Highlight cells C3 through C22, then press CTRL and D simultaneously (in Windows) or Open-Apple and D simultaneously (on a Macintosh). These both execute the command "copy down."
(2) Click on cell C3, and then put your cursor over the lower-right corner of this cell. In Windows, the big plus sign should become a small plus sign. On a Macintosh, the big plus sign becomes an open square. In either case, drag down your cursor to cell C22.

Using either procedure, the desired values (11 through 30) appear like magic! Click on cell C22. Note that the formula now reads = C21+1 Excel automatically "updates" the cell reference. The message, though, is still the same: "Take the number right above you, and add 1 to it." This is why the cell reference is called relative.

The formula requires that *T* be in K, so I make the required unit conversion in Column D. In cell D2, I type in =C2+273.15 and then execute the "copy down" procedure described above.

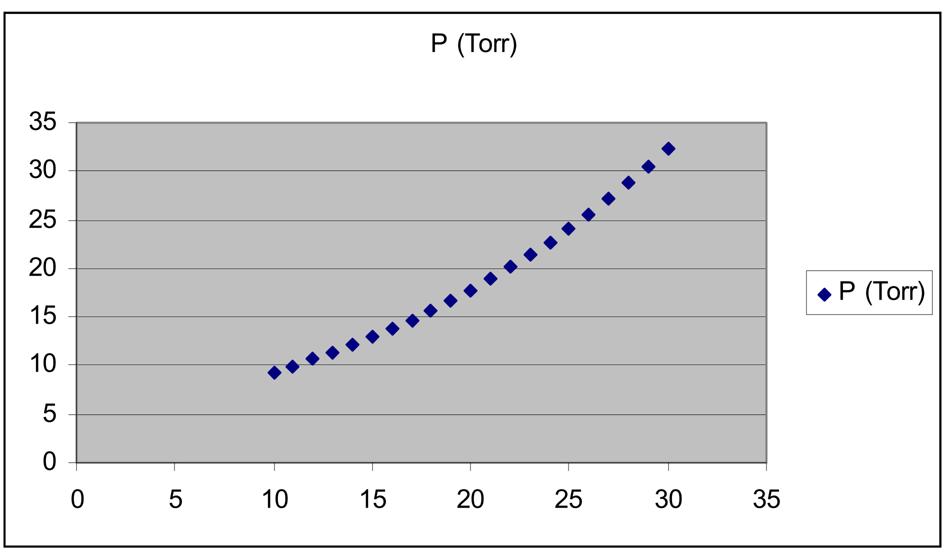
In Column E, I compute log P for each T in Column D. Here is the formula in cell E2: =$B$2-($B$3/(D2+$B$4)) The dollar signs ($) make cell references B2, B3, and B4 absolute—that is, they will not be automatically updated when we copy down in a moment. However, since I want the log P in each row to use the temperature in that row, I make the cell reference to temperature relative. (D2 in the formula really means, "Use the value in the cell just to the left of you.) Copy down as usual. Click on cell E22, and note the formula present: =$B$2-($B$3/(D22+$B$4)) The dollar signs indeed "lock in" references to the three cells containing constants. However, the temperature reference is updated to D22.

Excel can be a great time-saver. Instead of using your calculator to do unit conversions or other math procedures on a column of numbers, use neighboring columns in Excel to do them. For example, we need to go from the log (base 10) of vapor pressure to vapor pressure. We do that by typing into cell F2 =10^E2 and copy down to cell F22. Finally, to convert from atm to Torr, we type into cell G2 =F2*760 and copy down to cell G22.

Now, let's generate the plot! The general approach is to highlight the numbers to be plotted, then choose "Insert Chart" (using the Insert pull-down menu). (Instead of using the pull-down menu, you can simply press the chart button (labeled by a tri-color bar graph), if such a button is present on your tool bar.) However, note that we are required to plot T in degrees C on the x-axis, and P in Torr on the y-axis. After selecting one set of cells, you can select another set of cells in a non-adjacent column by keeping the CTRL key (Windows)/Open-Apple key (Macintosh) pressed down.

This will bring up the Chart Wizard. Choose Chart type XY (Scatter) in Step 1. Step 2 lets you redefine which cells are being plotted; we'll skip that here. Step 3 lets you tend to chart appearance: adding axis labels, etc. Doing this is **required**; you should always document your charts. Step 4 lets you decide if you want to paste the chart into your current worksheet (a good idea if you want to fit everything on one sheet of paper when you print out) or on a new sheet.

Note that once a chart has been created, you are free to continue altering its appearance. In particular, I highly recommend that you double-click on both the x-axis and the y-axis and reset the Minimum and Maximum values (found on the Scale tab) so that the points fill all available space on the graph (like the plot on p. 1). In other words, don't leave your graph like this:



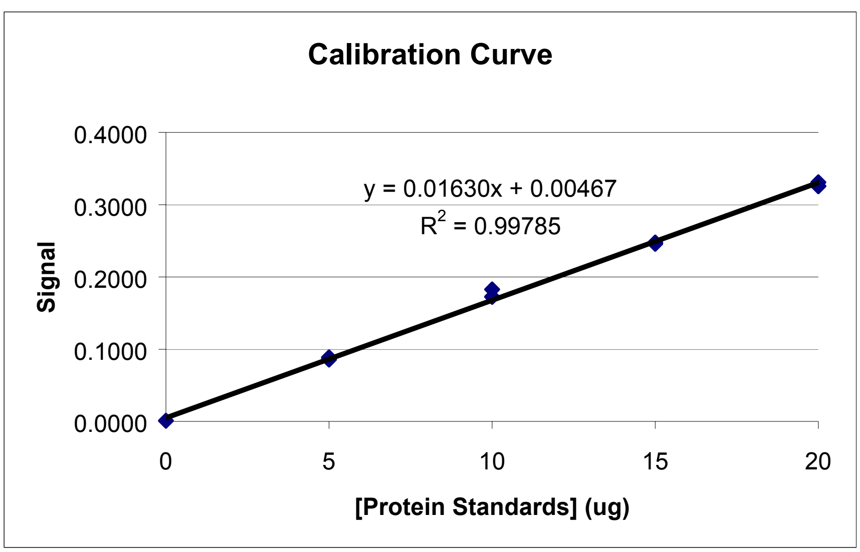
II. Constructing a Calibration Curve by the Method of Least Squares

A. First Iteration: Using Add Trendline

	A	B	C	D	E	F	G	H	I
1	Calibration Curve Data from Harris Table 5-2								
2	Protein (ug)	Signal	Corrected Signal						
3	0	0.099	-0.0003	Cell C3 has the formula =b3-c18					
4	0	0.099	-0.0003	(note the absolute cell reference to the mean blank signal)					
5	0	0.100	0.0007						
6	5	0.185	0.0857						
7	5	0.187	0.0877						
8	5	0.188	0.0887						
9	10	0.282	0.1827						
10	10	0.272	0.1727						
11	10	0.272	0.1727						
12	15	0.345	0.2457						
13	15	0.347	0.2477						
14	20	0.425	0.3257						
15	20	0.425	0.3257						
16	20	0.430	0.3307						
17									
18		Mean blank:	0.099333	C18 has the function =average(b3:b5)					
19				(note the use of the colon to indicate a range of cells)					

After you create the above spreadsheet, select the data in Columns A and C and generate a plot (as described on p. 3). Next, click on the points, and do the following:

- Select "Add Trendline" under the Chart pull-down menu.
- Under the "Type" tab, choose a linear Trend/Regression Type.
- Under the "Options" tab, choose to display both the equation and R-squared (R^2) value on the chart.
- Click on your trendline box and go to "Selected Data Labels" in the Format pull-down menu. Under the Number tab, choose to display at least three figures for your parameters.



The correlation coefficient R^2 is a good <u>qualitative</u> measure of linearity, but…

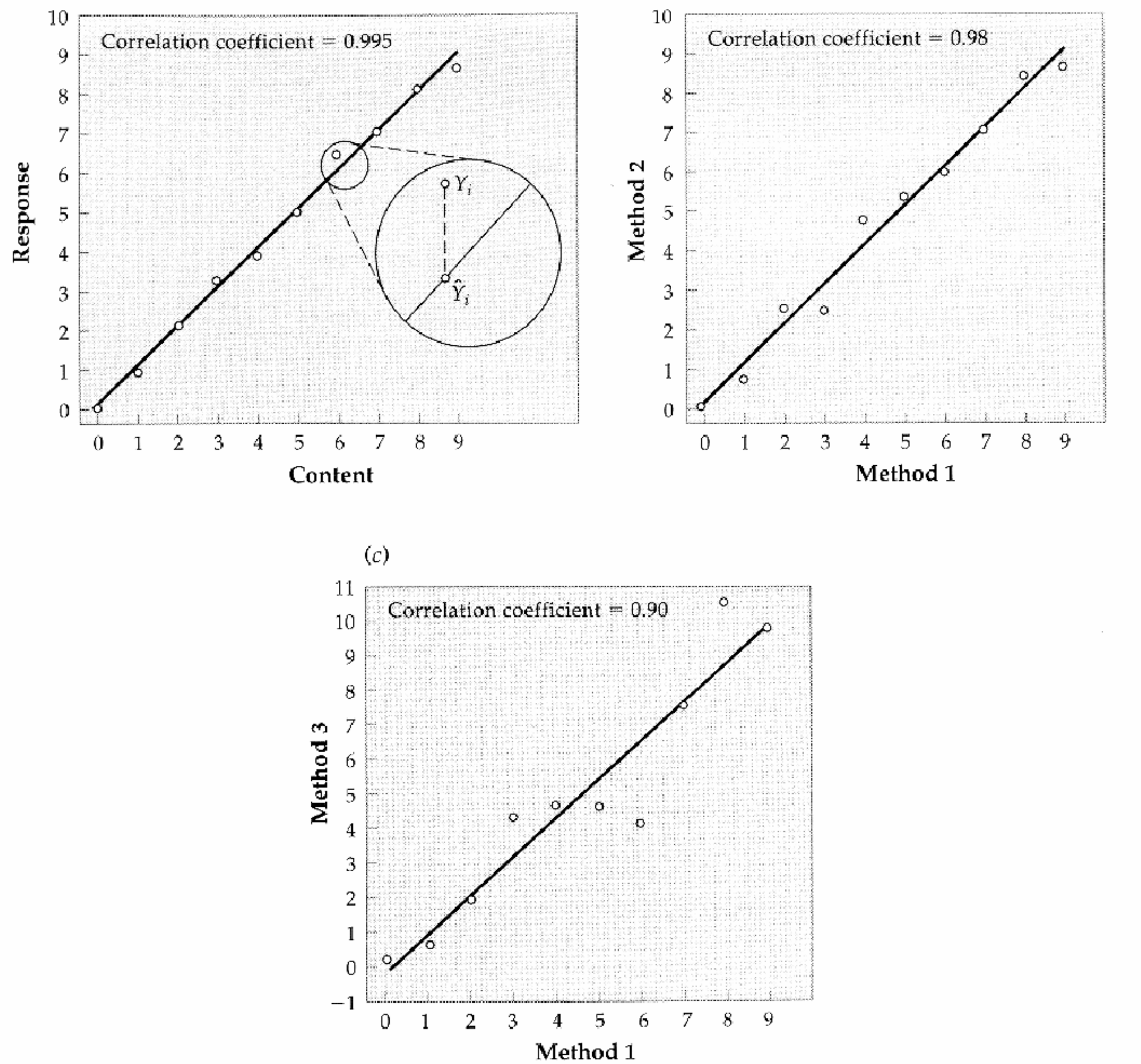


FIGURE 2.8 ▲
Example of linear regression.

(a) shows calibration data fitted with a least squares straight line. The correlation coefficient is 0.995. The magnifier shows the relationship between Y_i and $\hat{Y}_i$. They are at the same X_i-value.
(b) shows a validation between two methods, but with a correlation coefficient of 0.98.
(c) shows another validation with a much worse method that, for the data here, appears to be invalid in the middle of the range. Nevertheless, the correlation coefficient is 0.90.

from Rubinson, K. A.; Rubinson, J. F. *Contemporary Instrumental Analysis*; Prentice-Hall: Upper Saddle River, NJ, 2000; Chapter 2.

B. Second Iteration: Using the Excel Array Function LINEST

	A	B	C	D	E	F	G	H
1	Calibration Curve Data from Harris Table 5-2							
2	Protein (ug)	Signal	Corrected Signal					
3	0	0.099	-0.0003	Cell C3 has the formula =b3-c18				
4	0	0.099	-0.0003	(note the absolute cell reference to the mean blank signal)				
5	0	0.100	0.0007					
6	5	0.185	0.0857					
7	5	0.187	0.0877					
8	5	0.188	0.0887					
9	10	0.282	0.1827					
10	10	0.272	0.1727					
11	10	0.272	0.1727					
12	15	0.345	0.2457					
13	15	0.347	0.2477					
14	20	0.425	0.3257					
15	20	0.425	0.3257					
16	20	0.430	0.3307					
17								
18		Mean blank:	0.099333	C18 has the function =average(b3:b5)				
19								
20			Slope (m)	0.0162963	0.00466667	y-Intercept (b)		
21	Standard Error in Slope (s_m)			0.00021847	0.00262749	Standard Error in y-Intercept (s_b)		
22	Correlation Coefficient (R^2)			0.99784795	0.00587525	Standard Error in Signal Measurement (s_y)		
23				5564.07112	12	Degrees of Freedom		
24				0.19206349	0.00041422	(14 data - 1 for slope -1 for y-intercept)		

LINEST is an example of an array function with four arguments. In the above spreadsheet, you would enter it as follows:

- Select a 2-column by 5-row array of cells (D20:E24 above) (Note the use of a colon to specify a range of cells.)
- Type in `=linest(c3:c16,a3:a16,true,true)`

  LINEST's first argument is the range of cells containing y-values. The second argument is the range of cells containing x-values. (Excel will complain if the number of y-values does not match the number of x-values.) The third argument (true or false) refers to whether we want to optimize the y-intercept (true) or force the y-intercept to be zero (false). The fourth argument (true or false) is asking if we want other statistical parameters besides *m* and *b*. <u>Always say true for the last two arguments</u>.

- (On Windows machines:) Press CTRL-SHIFT-ENTER simultaneously
- (On Macintoshes:) Press OpenApple-SHIFT-ENTER simultaneously

The above spreadsheet labels seven of the ten parameters computed by LINEST. It reports not only the least squares parameters *m* and *b*, but also the standard errors of measurement in *m* (that is, s_m), in *b* (that is, s_b), and in a reading *y* made on a sample (that is, s_y). Because these are standard errors of measurement (that is, standard deviations divided by $\sqrt{n}$), you obtain 95% confidence intervals for *m*, *b*, and *y* simply by multiplying s_m, s_b, and s_y by the appropriate value of Student's *t* for ***n*-2** degrees of freedom. We lose **two** degrees of freedom since we have calculated both a slope and a y-intercept from the data. (Note that Harris is wrong: LINEST does not report standard deviations in *m*, *b*, and *y*: they have already been divided by $\sqrt{n}$.)

The standard error in the slope is enough information in many cases (such as in Physical Chemistry I experiments), but in Analytical Chemistry, we want to quantify the error in *x*, the concentration corresponding to a measurement *y*….

C. Final Iteration: Explicit Evaluation of the Least-Squares Formulas
(also see spreadsheet in Harris Figure 5-9)

	A	B	C	D	E	F	G	H
1	Calibration Curve Data from Harris Table 5-2							
2	x_i	Signal	y_i	x_iy_i	x_i^2	d_i	d_i^2	
3	0	0.099	-0.0003	0	0	-5.000E-03	2.500E-05	
4	0	0.099	-0.0003	0	0	-5.000E-03	2.500E-05	
5	0	0.100	0.0007	0	0	-4.000E-03	1.600E-05	
6	5	0.185	0.0857	0.4283	25	-4.815E-04	2.318E-07	
7	5	0.187	0.0877	0.4383	25	1.519E-03	2.306E-06	
8	5	0.188	0.0887	0.4433	25	2.519E-03	6.343E-06	
9	10	0.282	0.1827	1.8267	100	1.504E-02	2.261E-04	
10	10	0.272	0.1727	1.7267	100	5.037E-03	2.537E-05	
11	10	0.272	0.1727	1.7267	100	5.037E-03	2.537E-05	
12	15	0.345	0.2457	3.6850	225	-3.444E-03	1.186E-05	
13	15	0.347	0.2477	3.7150	225	-1.444E-03	2.086E-06	
14	20	0.425	0.3257	6.5133	400	-4.926E-03	2.426E-05	
15	20	0.425	0.3257	6.5133	400	-4.926E-03	2.426E-05	
16	20	0.430	0.3307	6.6133	400	7.407E-05	5.487E-09	
17	**135**		**2.2653**	**33.6300**	**2025**	**-7.633E-16**	**4.142E-04**	**sums**
18								
19	# of points:	14		Mean blank:	0.099			
20								
21	D =	10125		std err (y) =	0.005875		Measured y =	0.302
22	m =	0.0163		std err (m) =	0.000218		k =	1
23	b =	0.00467		std err (b) =	0.002627		Derived x =	18.25
24							std err (x) =	0.39

First, compute values of $x_i y_i$ and x_i^2 for each point, then sum up columns A, C, D, and E. Then evaluate the following formulas (note that *n* is the number of points):

$$D = n\sum x_i^2 - \left(\sum x_i\right)^2 \qquad m = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{D} \qquad b = \frac{\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i}{D}$$

Then use *m* and *b* to compute a deviation ($d_i = y_i - mx_i - b$) for each point (Column F), and its square (d_i^2) (Column G). Sum up Columns F and G. This equips us to compute the standard errors of measurement in a given signal measurement (*y*), slope, and intercept:

$$s_y = \sqrt{\frac{\sum d_i^2}{n-2}} \qquad s_m = s_y\sqrt{\frac{n}{D}} \qquad s_b = s_y\sqrt{\frac{\sum x_i^2}{D}}$$

And the payoff: For *k* measurements on an unknown, we get an average signal *y*. We solve for the unknown's concentration *x*. We then calculate the standard error of measurement in *x* thus:

$$s_x = \frac{s_y}{|m|}\sqrt{\frac{1}{k} + \frac{nx^2}{D} + \frac{\sum x_i^2}{D} - \frac{2x\sum x_i}{D}}$$

As before, you compute 95% confidence intervals by multiplying s_m, s_b, s_y, or s_x by the appropriate value of Student's *t* for ***n*-2** degrees of freedom.