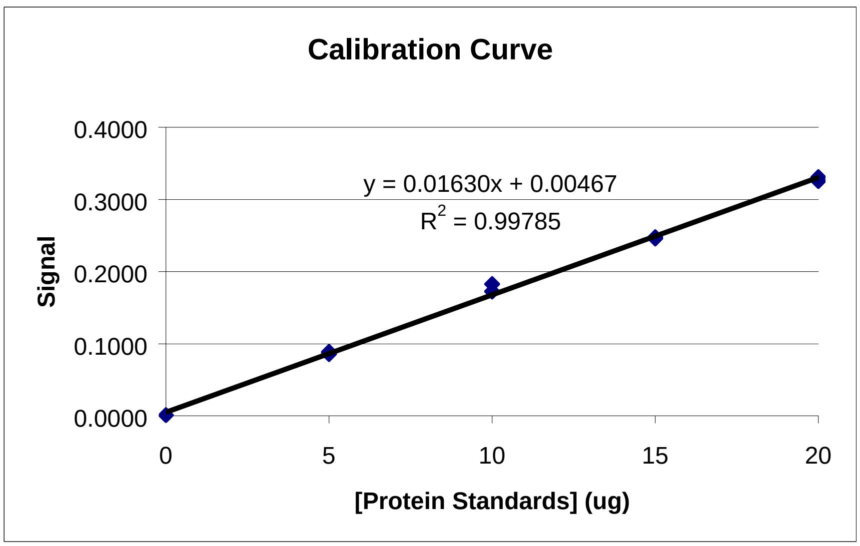
Analytical Chemistry
Constructing a Calibration Curve by the Method of Least Squares

A. <u>First Iteration: Using Add Trendline</u>

	A	B	C	D	E	F	G	H	I
1	Calibration Curve Data from Harris Table 4-7								
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. 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 qualitative measure of linearity, but…

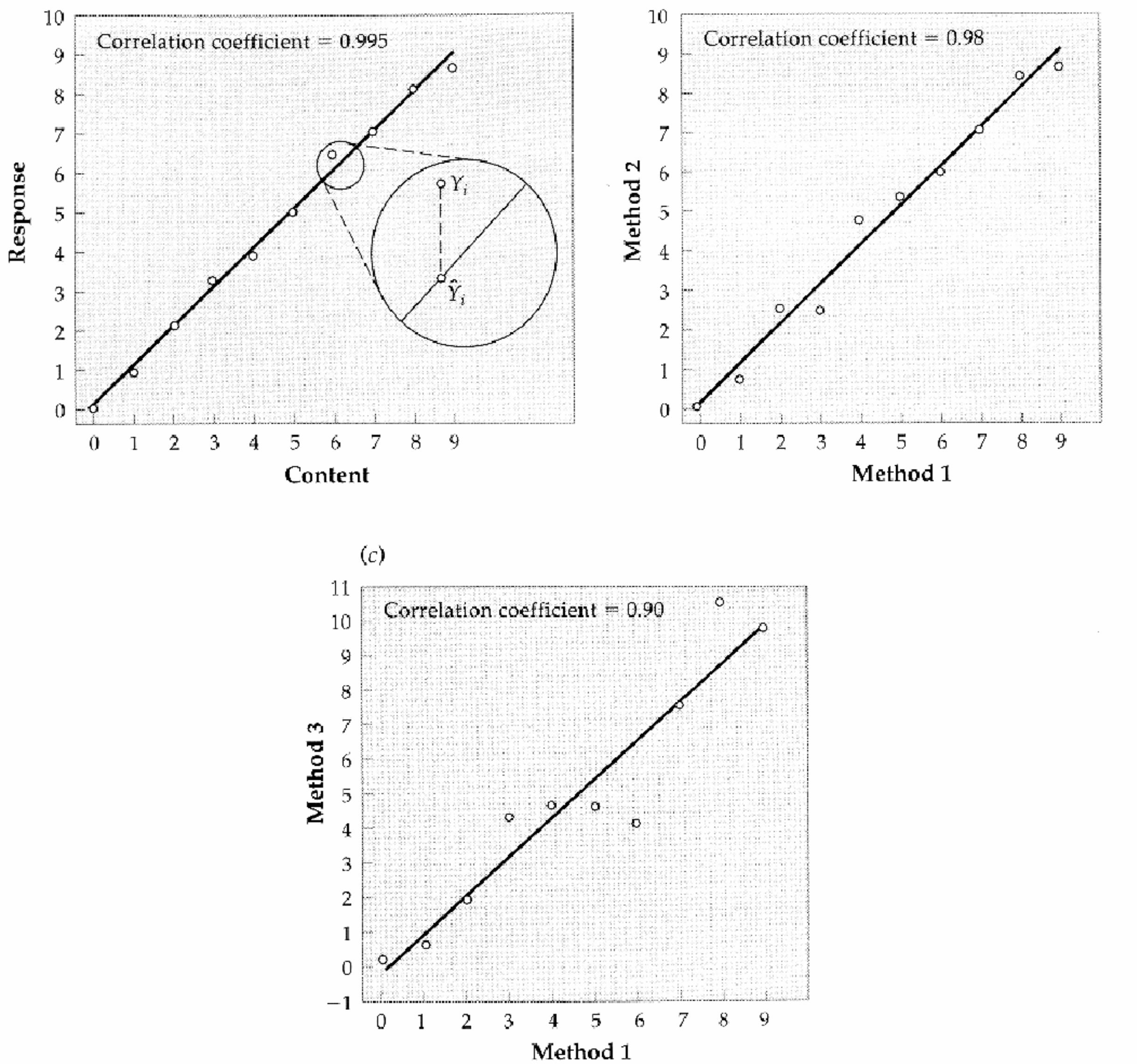


FIGURE 2.8 ▲
Example of linear regression.
(a) shows calibration data fitted with a least squares straight line. The correlation coefficient is 0.995. The magnifier shows the relationship between Y_i and $\hat{Y}_i$. They are at the same X_i-value.
(b) shows a validation between two methods, but with 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 4-7							
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*. Always say true for the last two arguments.

- (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: Treating the Correlation in the Errors in Slope and Y-Intercept

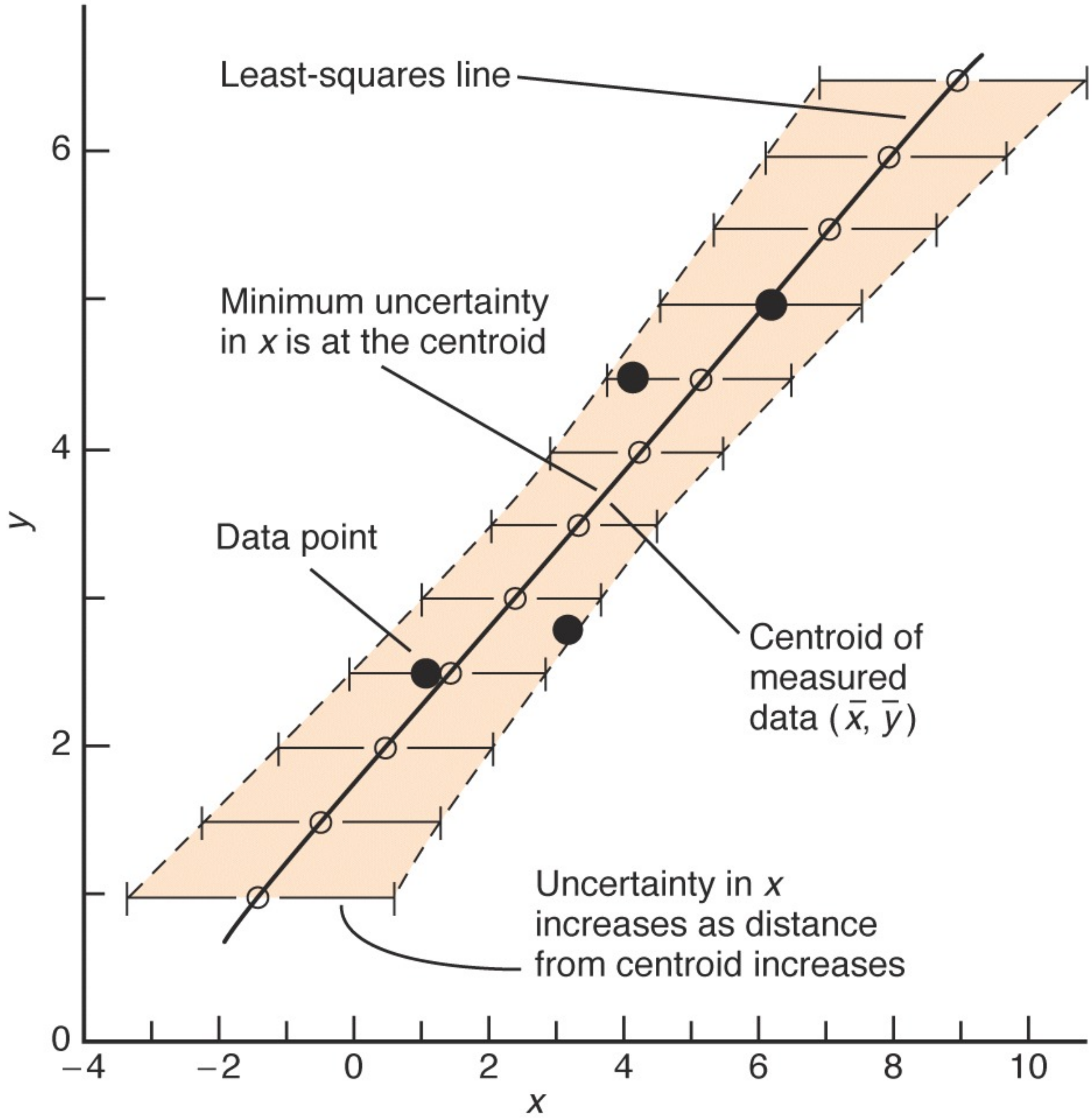
(also see spreadsheet in Harris Figure 4-13)

	A	B	C	D	E	F	G	H	I	J	K	L
1	Calibration Curve Data from Harris Table 4-7; Analysis in Harris Figure 4-13											
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	0.004666667	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	0.005875246	Standard Error in Signal Measurement (s_y)						
23				5564.07112	12	Degrees of Freedom						
24				0.19206349	0.000414222	(14 data - 1 for slope -1 for y-intercept)						
25												
26				n = number of pts on calibration curve =	14	E26 has the function =count(A3:A16)						
27				Mean y for pts on calibration curve =	0.16180952	E27 has the function =average(C3:C16)						
28				Sum of squares of deviations in x =	723.2143	E28 has the function =devsq(A3:A16)						
29												
30				Measured y (corrected) =	0.302							
31				k = Number of times y measured =	1							
32				Derived x =	18.25	E32 has the function =(E30-E20)/D20						
33				std err (x) =	0.39	E33 has the function =(E22/D20)*SQRT((1/E31)+(1/E26)+((E30-E27)^2)/(D20^2*E28))						

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:

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

As before, you compute 95% confidence intervals by multiplying s_x by the appropriate value of Student's t for ***n*-2** degrees of freedom. (Note that while a larger value of k increases the precision of our determination of x, it does not affect how many degrees of freedom we have.)



The shaded area shows the standard errors in x (s_x) computed correctly, that is, by treating the correlation in s_m and s_b. Note how the errors increase as one gets further away from the calibration curve's centroid.
(Taken from the 6th edition of Harris' *Quantitative Chemical Analysis*.)