

Chapter 5

Many of the problems in this chapter require a regression analysis. Although equations for these calculations are highlighted in the solution to the first such problem, for the remaining problems, both here and elsewhere in this text, the results of a regression analysis simply are provided. Be sure you have access to a scientific calculator, a spreadsheet program, such as Excel, or a statistical software program, such as R, and that you know how to use it to complete a regression analysis.

1. For each step in a dilution, the concentration of the new solution, C_{new} , is

$$C_{new} = \frac{C_{orig} V_{orig}}{V_{new}}$$

where C_{orig} is the concentration of the original solution, V_{orig} is the volume of the original solution taken, and V_{new} is the volume to which the original solution is diluted. A propagation of uncertainty for C_{new} shows that its relative uncertainty is

$$\frac{u_{C_{new}}}{C_{new}} = \sqrt{\left(\frac{u_{C_{orig}}}{C_{orig}}\right)^2 + \left(\frac{u_{V_{orig}}}{V_{orig}}\right)^2 + \left(\frac{u_{V_{new}}}{V_{new}}\right)^2}$$

For example, if we dilute 10.00 mL of the 0.1000 M stock solution to 100.0 mL, C_{new} is 1.000×10^{-2} M and the relative uncertainty in C_{new} is

$$\frac{u_{C_{new}}}{C_{new}} = \sqrt{\left(\frac{0.0002}{0.1000}\right)^2 + \left(\frac{0.02}{10.00}\right)^2 + \left(\frac{0.08}{100.0}\right)^2} = 2.94 \times 10^{-3}$$

The absolute uncertainty in C_{new} , therefore, is

$$u_{C_{new}} = (1.000 \times 10^{-2} \text{ M}) \times (2.94 \times 10^{-3}) = 2.94 \times 10^{-5} \text{ M}$$

The relative and the absolute uncertainties for each solution's concentration are gathered together in the tables that follow (all concentrations are given in mol/L and all volumes are given in mL). The uncertainties in the volumetric glassware are from Table 4.2 and Table 4.3. For a V_{orig} of 0.100 mL and of 0.0100 mL, the uncertainties are those for a 10–100 μL digital pipet.

For a serial dilution, each step uses a 10.00 mL volumetric pipet and a 100.0 mL volumetric flask; thus

C_{new}	C_{orig}	V_{orig}	V_{new}	$u_{V_{orig}}$	$u_{V_{new}}$
1.000×10^{-2}	0.1000	10.00	100.0	0.02	0.08
1.000×10^{-3}	1.000×10^{-2}	10.00	100.0	0.02	0.08
1.000×10^{-4}	1.000×10^{-3}	10.00	100.0	0.02	0.08
1.000×10^{-5}	1.000×10^{-4}	10.00	100.0	0.02	0.08

See Chapter 4C to review the propagation of uncertainty.

C_{new}	C_{orig}	$\frac{u_{C_{new}}}{C_{new}}$	$u_{C_{new}}$
1.000×10^{-2}	0.1000	2.94×10^{-3}	2.94×10^{-5}
1.000×10^{-3}	1.000×10^{-2}	3.64×10^{-3}	3.64×10^{-6}
1.000×10^{-4}	1.000×10^{-3}	4.23×10^{-3}	4.23×10^{-7}
1.000×10^{-5}	1.000×10^{-4}	4.75×10^{-3}	4.75×10^{-8}

For the set of one-step dilutions using the original stock solution, each solution requires a different volumetric pipet; thus

C_{new}	C_{orig}	V_{orig}	V_{new}	$u_{V_{orig}}$	$u_{V_{new}}$
1.000×10^{-2}	0.1000	10.00	100.0	0.02	0.08
1.000×10^{-3}	0.1000	1.000	100.0	0.006	0.08
1.000×10^{-4}	0.1000	0.100	100.0	8.00×10^{-4}	0.08
1.000×10^{-5}	0.1000	0.0100	100.0	3.00×10^{-4}	0.08

C_{new}	C_{orig}	$\frac{u_{C_{new}}}{C_{new}}$	$u_{C_{new}}$
1.000×10^{-2}	0.1000	2.94×10^{-3}	2.94×10^{-5}
1.000×10^{-3}	0.1000	6.37×10^{-3}	6.37×10^{-6}
1.000×10^{-4}	0.1000	8.28×10^{-3}	8.28×10^{-7}
1.000×10^{-5}	0.1000	3.01×10^{-2}	3.01×10^{-7}

Note that for each C_{new} , the absolute uncertainty when using a serial dilution always is equal to or better than the absolute uncertainty when using a single dilution of the original stock solution. More specifically, for a C_{new} of 1.000×10^{-3} M and of 1.000×10^{-4} M, the improvement in the absolute uncertainty is approximately a factor of 2, and for a C_{new} of 1.000×10^{-5} M, the improvement in the absolute uncertainty is approximately a factor of 6. This is a distinct advantage of a serial dilution. On the other hand, for a serial dilution a determinate error in the preparation of the 1.000×10^{-2} M solution carries over as a determinate error in each successive solution, which is a distinct disadvantage.

2. We begin by determining the value for k_A in the equation

$$S_{total} = k_A C_A + S_{reag}$$

where S_{total} is the average of the three signals for the standard of concentration C_A , and S_{reag} is the signal for the reagent blank. Making appropriate substitutions

$$0.1603 = k_A(10.0 \text{ ppm}) + 0.002$$

and solving for k_A gives its value as 0.01583 ppm^{-1} . Substituting in the signal for the sample

$$0.118 = (0.01583 \text{ ppm}^{-1}) C_A + 0.002$$

and solving for C_A gives the analyte's concentration as 7.33 ppm.

3. This standard addition follows the format of equation 5.9

$$\frac{S_{\text{samp}}}{C_A \frac{V_o}{V_f}} = \frac{S_{\text{spike}}}{C_A \frac{V_o}{V_f} + C_{\text{std}} \frac{V_{\text{std}}}{V_f}}$$

in which both the sample and the standard addition are diluted to the same final volume. Making appropriate substitutions

$$\frac{0.235}{C_A \times \frac{10.00 \text{ mL}}{25.00 \text{ mL}}} = \frac{0.502}{C_A \times \frac{10.00 \text{ mL}}{25.00 \text{ mL}} + (1.00 \text{ ppm}) \times \frac{10.00 \text{ mL}}{25.00 \text{ mL}}}$$

$$0.0940 C_A + 0.0940 \text{ ppm} = 0.2008 C_A$$

and solving gives the analyte's concentration, C_A , as 0.800 ppm. The concentration of analyte in the original solid sample is

$$\frac{(0.880 \text{ mg/L})(0.250 \text{ L}) \left(\frac{1 \text{ g}}{1000 \text{ mg}} \right)}{10.00 \text{ g sample}} \times 100 = 2.20 \times 10^{-3} \% \text{ w/w}$$

4. This standard addition follows the format of equation 5.11

$$\frac{S_{\text{samp}}}{C_A} = \frac{S_{\text{spike}}}{C_A \frac{V_o}{V_o + V_{\text{std}}} + C_{\text{std}} \frac{V_{\text{std}}}{V_o + V_{\text{std}}}}$$

in which the standard addition is made directly to the solution that contains the analyte. Making appropriate substitutions

$$\frac{11.5}{C_A} = \frac{23.1}{C_A \frac{50.00 \text{ mL}}{50.00 \text{ mL} + 1.00 \text{ mL}} + \frac{(10.0 \text{ ppm})(1.00 \text{ mL})}{50.00 \text{ mL} + 1.00 \text{ mL}}}$$

$$23.1 C_A = 11.27 C_A + 2.255 \text{ ppm}$$

and solving gives the analyte's concentration, C_A , as 0.191 ppm.

5. To derive a standard additions calibration curve using equation 5.10

$$S_{\text{spike}} = k_A \left(C_A \frac{V_o}{V_o + V_{\text{std}}} + C_{\text{std}} \frac{V_{\text{std}}}{V_o + V_{\text{std}}} \right)$$

we multiply through both sides of the equation by $V_o + V_{\text{std}}$

$$S_{\text{spike}} (V_o + V_{\text{std}}) = k_A C_A V_o + k_A C_{\text{std}} V_{\text{std}}$$

As shown in Figure SM5.1, the slope is equal to k_A and the y -intercept is equal to $k_A C_A V_o$. The x -intercept occurs when $S_{\text{spike}} (V_o + V_{\text{std}})$ equals zero; thus

$$0 = k_A C_A V_o + k_A C_{\text{std}} V_{\text{std}}$$

Here we assume that a part per million is equivalent to mg/L.

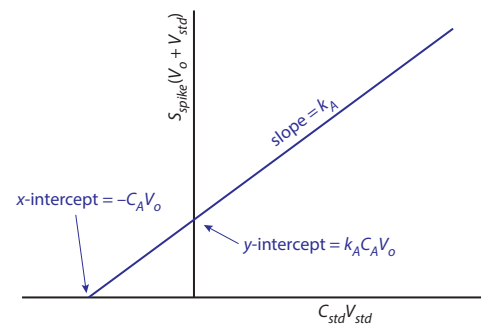


Figure SM5.1 Standard additions calibration curve based on equation 5.10.

and the x -intercept is equal to $-C_A V_o$. We must plot the calibration curve this way because if we plot S_{spike} on the y -axis versus $C_{std} \times \{V_{std} / (V_o + V_{std})\}$ on the x -axis, then the term we identify as y -intercept

$$\frac{k_A C_A V_o}{V_o + V_{std}}$$

is not a constant because it includes a variable, V_{std} , whose value changes with each standard addition.

6. Because the concentration of the internal standard is maintained at a constant level for both the sample and the standard, we can fold the internal standard's concentration into the proportionality constant K in equation 5.12; thus, using S_A , S_{IS} , and C_A for the standard

$$\frac{S_A}{S_{IS}} = \frac{0.155}{0.233} = \frac{k_A C_A}{k_{IS} C_{IS}} = K C_A = K(10.00 \text{ mg/L})$$

gives K as 0.06652 L/mg. Substituting in S_A , S_{IS} , and K for the sample

$$\frac{0.155}{0.233} = (0.06652 \text{ L/mg}) C_A$$

gives the concentration of analyte in the sample as 20.8 mg/L.

7. For each pair of calibration curves, we seek to find the calibration curve that yields the smallest uncertainty as expressed in the standard deviation about the regression, s_r , the standard deviation in the slope, s_{b_1} , or the standard deviation in the y -intercept, s_{b_0} .

(a) The calibration curve on the right is the better choice because it uses more standards. All else being equal, the larger the value of n , the smaller the value for s_r in equation 5.19, and for s_{b_0} in equation 5.21.

(b) The calibration curve on the left is the better choice because the standards are more evenly spaced, which minimizes the term $\sum x_i^2$ in equation 5.21 for s_{b_0} .

(c) The calibration curve on the left is the better choice because the standards span a wider range of concentrations, which minimizes the term $\sum (x_i - \bar{X})^2$ in equation 5.20 and in equation 5.21 for s_{b_1} and s_{b_0} , respectively.

8. To determine the slope and the y -intercept for the calibration curve at a pH of 4.6 we first need to calculate the summation terms that appear in equation 5.17 and in equation 5.18; these are:

$$\begin{aligned} \sum x_i &= 308.4 & \sum y_i &= 131.0 \\ \sum x_i y_i &= 8397.5 & \sum x_i^2 &= 19339.6 \end{aligned}$$

Substituting these values into the equation 5.17

$$b_1 = \frac{(6 \times 8397.5) - (308.4 \times 131.0)}{(6 \times 19339.6) - (308.4)^2} = 0.477$$

As a reminder, for this problem we will work through the details of an unweighted linear regression calculation using the equations from the text. For the remaining problems, it is assumed you have access to a calculator, a spreadsheet, or a statistical program that can handle most or all of the relevant calculations for an unweighted linear regression.

gives the slope as 0.477 nA/nM, and substituting into equation 5.18

$$b_0 = \frac{131.0 - (0.477 \times 308.4)}{6} = -2.69$$

gives the y -intercept as -2.69 nA. The equation for the calibration curve is

$$S_{total} = 0.477 \text{ nA/nM} \times C_{Cd} - 2.69 \text{ nA}$$

Figure SM5.2 shows the calibration data and the calibration curve.

To find the confidence intervals for the slope and for the y -intercept, we use equation 5.19 to calculate the standard deviation about the regression, s_r , and use equation 5.20 and equation 5.21 to calculate the standard deviation in the slope, s_{b_1} , and the standard deviation in the y -intercept, s_{b_0} , respectively. To calculate s_r we first calculate the predicted values for the signal, \hat{y}_i , using the known concentrations of Cd^{2+} and the regression equation, and the squared residual errors, $(y_i - \hat{y}_i)^2$; the table below summarizes these results

x_i	y_i	\hat{y}_i	$(y_i - \hat{y}_i)^2$
15.4	4.8	4.66	0.0203
30.4	11.4	11.81	0.7115
44.9	18.2	18.73	0.2382
59.0	26.6	25.46	1.3012
72.7	32.3	32.00	0.0926
86.0	37.7	38.34	0.4110

Adding together the last column, which equals 2.2798, gives the numerator for equation 5.19; thus, the standard deviation about the regression is

$$s_r = \sqrt{\frac{2.2798}{6-2}} = 0.7550$$

To calculate the standard deviations in the slope and in the y -intercept, we use equation 5.20 and equation 5.21, respectively, using the standard deviation about the regression and the summation terms outlined earlier; thus

$$s_{b_1} = \sqrt{\frac{6 \times (0.7550)^2}{(6 \times 19339.6) - (308.4)^2}} = 0.02278$$

$$s_{b_0} = \sqrt{\frac{(0.7550)^2 \times 19339.6}{(6 \times 19339.6) - (308.4)^2}} = 0.7258$$

With four degrees of freedom, the confidence intervals for the slope and the y -intercept are

$$\begin{aligned} \beta_1 &= b_1 \pm t_{s_{b_1}} = 0.477 \pm (2.776)(0.0128) \\ &= 0.477 \pm 0.036 \text{ nA/nM} \end{aligned}$$

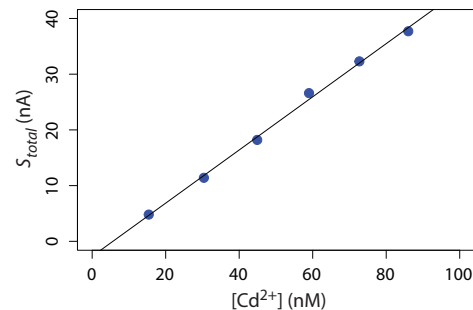


Figure SM5.2 Calibration curve at pH 4.6 for the data in Problem 5.8.

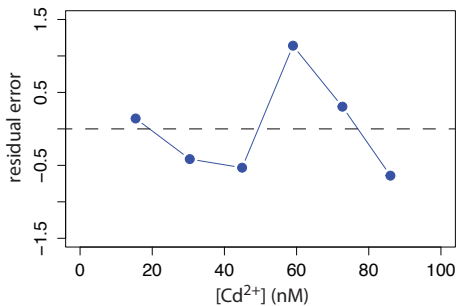


Figure SM5.3 Plot of the residual errors for the calibration standards in Problem 5.8 at a pH of 4.6.

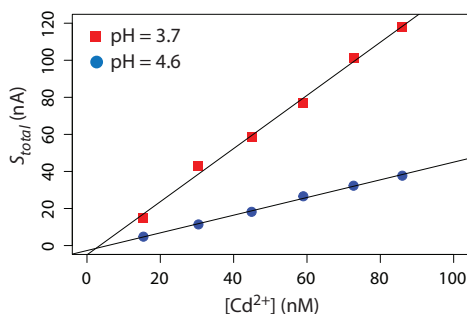


Figure SM5.4 Calibration curves for the data in Problem 5.8 at a **pH of 3.7** and at a **pH of 4.6**.

$$\begin{aligned}\beta_1 &= b_0 \pm ts_{b_0} = -2.69 \pm (2.776)(0.7258) \\ &= -2.69 \pm 2.01 \text{ nA}\end{aligned}$$

(b) The table below shows the residual errors for each concentration of Cd^{2+} . A plot of the residual errors (Figure SM5.3) shows no discernible trend that might cause us to question the validity of the calibration equation.

x_i	y_i	\hat{y}_i	$y_i - \hat{y}_i$
15.4	4.8	4.66	0.14
30.4	11.4	11.81	-0.41
44.9	18.2	18.73	-0.53
59.0	26.6	25.46	1.14
72.7	32.3	32.00	0.30
86.0	37.7	38.34	-0.64

(c) A regression analysis for the data at a pH of 3.7 gives the calibration curve's equation as

$$S_{\text{total}} = 1.43 \text{ nA/nM} \times C_{\text{Cd}} - 5.02 \text{ nA}$$

The more sensitive the method, the steeper the slope of the calibration curve, which, as shown in Figure SM5.4, is the case for the calibration curve at pH 3.7. The relative sensitivities for the two pHs is the ratio of their respective slopes

$$\frac{k_{\text{pH}3.7}}{k_{\text{pH}4.6}} = \frac{1.43}{0.477} = 3.00$$

The sensitivity at a pH of 3.7, therefore, is three times more sensitive than that at a pH of 4.6.

(d) Using the calibration curve at a pH of 3.7, the concentration of Cd^{2+} in the sample is

$$[\text{Cd}^{2+}] = \frac{S_{\text{total}} - b_0}{b_1} = \frac{66.3 \text{ nA} - (-5.02 \text{ nA})}{1.43 \text{ nA/nM}} = 49.9 \text{ nM}$$

To calculate the 95% confidence interval, we first use equation 5.25

$$s_{C_{\text{Cd}}} = \frac{s_r}{b_1} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(\bar{S}_{\text{samp}} - \bar{S}_{\text{std}})^2}{(b_1)^2 \sum_{i=1}^n (C_{\text{std}_i} - \bar{C}_{\text{std}})^2}}$$

to determine the standard deviation in the concentration where the number of samples, m , is one, the number of standards, n , is six, the standard deviation about the regression, s_r , is 2.826, the slope, b_1 , is 1.43, the average signal for the one sample, \bar{S}_{samp} , is 66.3, and the average signal for the six standards, \bar{S}_{std} , is 68.7. At first glance, the term $\sum (C_{\text{std}_i} - \bar{C}_{\text{std}})^2$, where C_{std_i} is the concentration of the i^{th} standard and \bar{C}_{std} is the average concentration for the n standards, seems

cumbersome to calculate. We can simplify the calculation, however, by recognizing that $\sum (C_{std_i} - \bar{C}_{std})^2$ is the numerator in the equation that gives the standard deviation for the concentrations of the standards, s_{Cd} . Because s_{Cd} is easy to determine using a calculator, a spreadsheet, or a statistical software program, it is easy to calculate $\sum (C_{std_i} - \bar{C}_{std})^2$; thus

$$\sum_{i=1}^n (C_{std_i} - \bar{C}_{std})^2 = (n - 1)(s_{Cd})^2 = (6 - 1)(26.41)^2 = 3487$$

Substituting all terms back into equation 5.25 gives the standard deviation in the concentration as

$$s_{Cca} = \frac{2.826}{1.43} \sqrt{\frac{1}{1} + \frac{1}{6} + \frac{(66.3 - 68.7)^2}{(1.43)^2(3487)}} = 2.14$$

The 95% confidence interval for the sample's concentration, therefore, is

$$\mu_{Cd} = 49.9 \pm (2.776)(2.14) = 49.9 \pm 5.9 \text{ nM}$$

9. The standard addition for this problem follows equation 5.10, which, as we saw in Problem 5.5, is best treated by plotting $S_{spike}(V_o + V_{std})$ on the y -axis vs. $C_s V_s$ on the x -axis, the values for which are

V_{std} (mL)	S_{spike} (arb. units)	$S_{spike}(V_o + V_{std})$	$C_{std}V_{std}$
0.00	0.119	0.595	0.0
0.10	0.231	1.178	60.0
0.20	0.339	1.763	120.0
0.30	0.442	2.343	180.0

Figure SM5.5 shows the resulting calibration curve for which the calibration equation is

$$S_{spike}(V_o + V_{std}) = 0.5955 + 0.009713 \times C_{std} V_{std}$$

To find the analyte's concentration, C_A , we use the absolute value of the x -intercept, $-C_A V_o$, which is equivalent to the y -intercept divided by the slope; thus

$$C_A V_o = C_A(5.00 \text{ mL}) = \frac{b_0}{k_A} = \frac{0.5955}{0.009713} = 61.31$$

which gives C_A as 12.3 ppb.

To find the 95% confidence interval for C_A , we use a modified form of equation 5.25 to calculate the standard deviation in the x -intercept

$$s_{C_A V_o} = \frac{s_r}{b_1} \sqrt{\frac{1}{n} + \frac{\{S_{spike}(V_o + V_{std})\}^2}{(b_1)^2 \sum_{i=1}^n (C_{std_i} V_{std_i} - \bar{C}_{std} V_{std})^2}}$$

where the number of standards, n , is four, the standard deviation about the regression, s_r , is 0.00155, the slope, b_1 , is 0.009713, the

$$s_{Cd} = \sqrt{\frac{\sum (C_{std_i} - \bar{C}_{std})^2}{n - 1}}$$

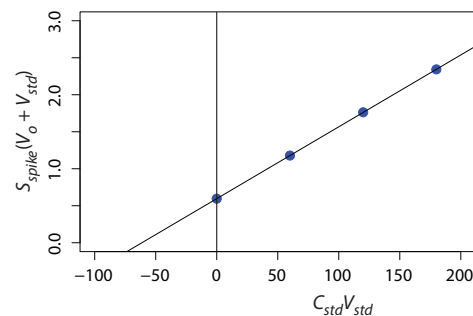


Figure SM5.5 Standard additions calibration curve for Problem 5.9.

average signal for the four standards, $\overline{S_{spike}(V_o + V_{std})}$, is 1.47, and the term $\sum (C_{std_i} V_{std_i} - \overline{C_{std} V_{std}})^2$ is 1.80×10^4 . Substituting back into this equation gives the standard deviation of the x -intercept as

$$s_{C_A V_o} = \frac{0.00155}{0.009713} \sqrt{\frac{1}{4} + \frac{\{1.47\}^2}{(0.009713)^2 (1.8 \times 10^4)}} = 0.197$$

Dividing $s_{C_A V_o}$ by V_o gives the standard deviation in the concentration, s_{C_A} , as

$$s_{C_A} = \frac{s_{C_A V_o}}{V_o} = \frac{0.197}{5.00} = 0.0393$$

The 95% confidence interval for the sample's concentration, therefore, is

$$\mu = 12.3 \pm (4.303)(0.0393) = 12.3 \pm 0.2 \text{ ppb}$$

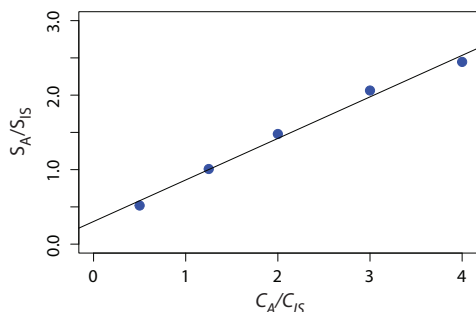


Figure SM5.6 Internal standards calibration curve for the data in Problem 5.10.

10. (a) For an internal standardization, the calibration curve places the signal ratio, S_A/S_{IS} , on the y -axis and the concentration ratio, C_A/C_{IS} , on the x -axis. Figure SM5.6 shows the resulting calibration curve, which is characterized by the following values

slope (b_1): 0.5576

y -intercept (b_0): 0.3037

standard deviation for slope (s_{b_1}): 0.0314

standard deviation for y -intercept (s_{b_0}): 0.0781

Based on these values, the 95% confidence intervals for the slope and the y -intercept are, respectively

$$\beta_0 = b_0 \pm t s_{b_0} = 0.3037 \pm (3.182)(0.0781) = 0.3037 \pm 0.2484$$

$$\beta_1 = b_1 \pm t s_{b_1} = 0.5576 \pm (3.182)(0.0314) = 0.5576 \pm 0.1001$$

(b) The authors concluded that the calibration model is inappropriate because the 95% confidence interval for the y -intercept does not include the expected value of 0.00. A close observation of Figure SM5.6 shows that the calibration curve has a subtle, but distinct curvature, which suggests that a straight-line is not a suitable model for this data.

11. Figure SM5.7 shows a plot of the measured values on the y -axis and the expected values on the x -axis, along with the regression line, which is characterized by the following values:

slope (b_1): 0.9996

y -intercept (b_0): 0.000761

standard deviation for slope (s_{b_1}): 0.00116

standard deviation for y -intercept (s_{b_0}): 0.00112

For the y -intercept, t_{exp} is

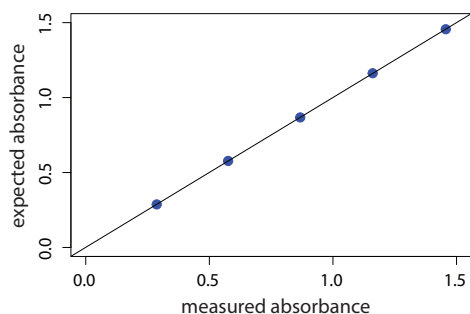


Figure SM5.7 Plot of the measured absorbance values for a series of spectrophotometric standards versus their expected absorbance values. The original data is from Problem 4.25.

$$t_{\text{exp}} = \frac{|\beta_0 - b_0|}{s_{b_0}} = \frac{|0.00 - 0.00761|}{0.00112} = 0.679$$

and t_{exp} for the slope is

$$t_{\text{exp}} = \frac{|\beta_1 - b_1|}{s_{b_1}} = \frac{|1.00 - 0.9996|}{0.00116} = 0.345$$

For both the y -intercept and the slope, t_{exp} is less than the critical value of $t(0.05,3)$, which is 3.182; thus, we retain the null hypothesis and have no evidence at $\alpha = 0.05$ that the y -intercept or the slope differ significantly from their expected values of zero, and, therefore, no evidence at $\alpha = 0.05$ that there is a difference between the measured absorbance values and the expected absorbance values.

12. (a) Knowing that all three data sets have identical regression statistics suggests that the three data sets are similar to each other. A close look at the values of y suggests that all three data sets show a general increase in the value of y as the value of x becomes larger, although the trend seems noisy.

(b) The results of a regression analysis are gathered here

parameter	Data Set 1	Data Set 2	Data Set 3
b_0	3.0001	3.0010	3.0025
b_1	0.5001	0.5000	0.4997
s_{b_0}	1.1247	1.1250	1.1245
s_{b_1}	0.1179	0.1180	0.1179
s_r	1.237	1.237	1.236

and are in agreement with the values reported in part (a). Figure SM5.8 shows the residual plots for all three data sets. For the first data set, the residual errors are scattered at random around a residual error of zero and show no particular trend, suggesting that the regression model provides a reasonable explanation for the data. For data set 2 and for data set 3, the clear pattern to the residual errors indicates that neither regression models is appropriate.

(c) Figure SM5.9 shows each data set with its regression line. For data set 1, the regression line provides a good fit to what is rather noisy data. For the second data set, we see that the relationship between x and y is not a straight-line and that a quadratic model likely is more appropriate. With the exception of an apparent outlier, data set 3 is a straight-line; removing the outlier is likely to improve the regression analysis.

(d) The apparent outlier is the third point in the data set ($x = 13.00$, $y = 12.74$). Figure SM5.10 shows the resulting regression line, for which

slope (b_1): 0.345

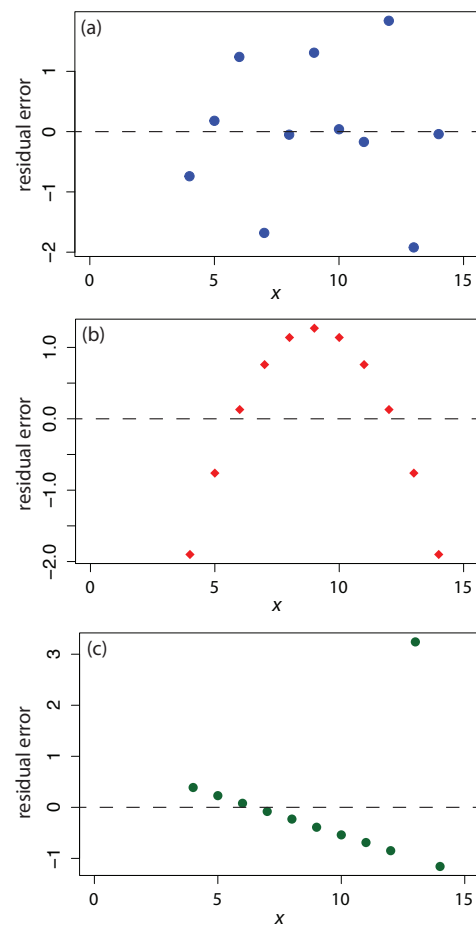


Figure SM5.8 Residual plots for (a) **data set 1**; (b) **data set 2**; and (c) **data set 3**. The dashed line in each plot shows the expected trend for the residual errors when the regression model provides a good fit to the data.

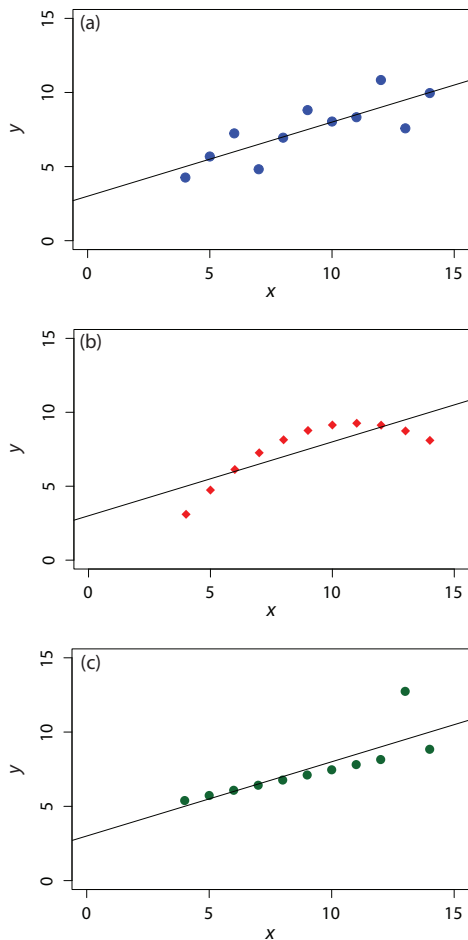


Figure SM5.9 Regression plots for the data from (a) **data set 1**; (b) **data set 2**; and (c) **data set 3**.

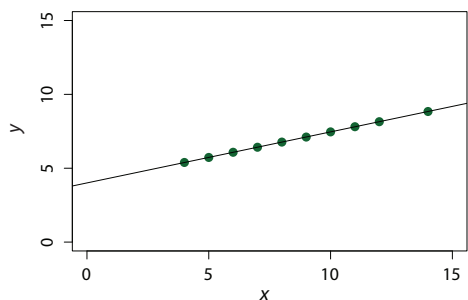


Figure SM5.10 Regression plot for data set 3 after removing the apparent outlier.

y -intercept (b_0): 4.01

standard deviation for slope (s_{b_1}): 0.00321

standard deviation for y -intercept (s_{b_0}): 0.00292

standard deviation about the regression (s_r): 0.00308

Note that s_r , s_{b_0} , and s_{b_1} are much smaller after we remove the apparent outlier, which is consistent with the better fit of the regression line to the data.

(e) The analysis of this data set drives home the importance of examining your data in a graphical form. As suggested earlier in the answer to part (a), it is difficult to see the underlying pattern in a data set when we look at numbers only.

13. To complete a weighted linear regression we first must determine the weighting factors for each concentration of thallium; thus

x_i	y_i (avg)	s_{y_i}	$(s_{y_i})^{-2}$	w_i
0.000	2.626	0.1137	77.3533	3.3397
0.387	8.160	0.2969	11.3443	0.4898
1.851	29.114	0.5566	3.2279	0.1394
5.734	85.714	1.1768	0.7221	0.0312

where y_i (avg) is the average of the seven replicate measurements for each of the i standard additions, and s_{y_i} is the standard deviation for these replicate measurements; note that the increase in s_{y_i} with larger values of x_i indicates that the indeterminate errors affecting the signal are not independent of the concentration of thallium, which is why a weighted linear regression is used here. The weights in the last column are calculated using equation 5.28 and, as expected, the sum of the weights is equal to the number of standards.

To calculate the y -intercept and the slope, we use equation 5.26 and equation 5.27, respectively, using the table below to organize the various summations

x_i	y_i (avg)	$w_i x_i$	$w_i y_i$	$w_i x_i^2$	$w_i x_i y_i$
0.000	2.626	0.0000	8.7701	0.0000	0.0000
0.387	8.160	0.1896	3.9968	0.0734	1.5467
1.851	29.114	0.2580	4.0585	0.4776	7.5123
5.734	85.714	0.1789	2.6743	1.0258	15.3343
totals		0.6265	19.4997	1.5768	24.3933

$$\begin{aligned}
 b_1 &= \frac{n \sum_{i=1}^n w_i x_i y_i - \sum_{i=1}^n w_i x_i \sum_{i=1}^n w_i y_i}{n \sum_{i=1}^n w_i x_i^2 - \left(\sum_{i=1}^n w_i x_i \right)^2} \\
 &= \frac{(4)(24.3933) - (0.6265)(19.4997)}{(4)(1.5768) - (0.6265)^2} = 14.43 \\
 b_0 &= \frac{\sum_{i=1}^n w_i y_i - b_1 \sum_{i=1}^n w_i x_i}{n} \\
 &= \frac{19.4997 - (14.431)(0.6265)}{4} = 2.61
 \end{aligned}$$

The calibration curve, therefore, is

$$S_{\text{total}} = 2.61 \mu\text{A} + (14.43 \mu\text{A}/\text{ppm}) \times C_{\text{Tl}}$$

Figure SM5.11 shows the calibration data and the weighted linear regression line.

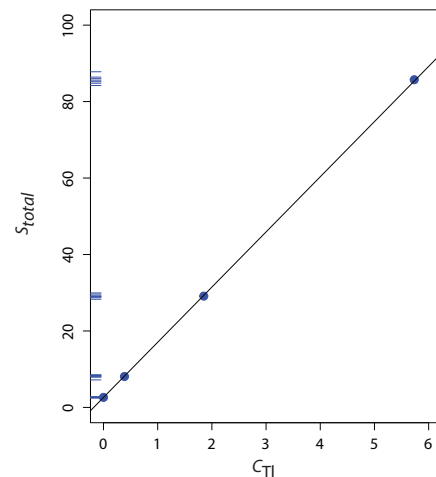


Figure SM5.11 Calibration data and calibration curve for the data in Problem 5.13. The individual points show the average signal for each standard and the calibration curve is from a weighted linear regression. The **blue** tick marks along the y -axis show the replicate signals for each standard; note that the spacing of these marks reflect the increased magnitude of the signal's indeterminate error for higher concentrations of thallium.

