Chapter 5

Standardizing Analytical Methods

Chapter Overview

The American Chemical Society's Committee on Environmental Improvement defines standardization as the process of determining the relationship between the signal and the amount of analyte in a sample.[1] In Chapter 3 we defined this relationship as

$$S_{total} = k_A n_A + S_{reag} \quad \text{or} \quad S_{total} = k_A C_A + S_{reag}$$

where S_{total} is the signal, n_A is the moles of analyte, C_A is the analyte's concentration, k_A is the method's sensitivity for the analyte, and S_{reag} is the contribution to S_{total} from sources other than the sample. To standardize a method we must determine values for k_A and S_{reag}. Strategies for accomplishing this are the subject of this chapter.

1 ACS Committee on Environmental Improvement "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," *Anal. Chem.* **1980**, *52*, 2242–2249.

5A Analytical Standards

To standardize an analytical method we use standards containing known amounts of analyte. The accuracy of a standardization, therefore, depends on the quality of the reagents and glassware used to prepare these standards. For example, in an acid–base titration the stoichiometry of the acid–base reaction defines the relationship between the moles of analyte and the moles of titrant. In turn, the moles of titrant is the product of the titrant's concentration and the volume of titrant needed to reach the equivalence point. The accuracy of a titrimetric analysis, therefore, can be no better than the accuracy to which we know the titrant's concentration.

See Chapter 9 for a thorough discussion of titrimetric methods of analysis.

5A.1 Primary and Secondary Standards

We divide analytical standards into two categories: primary standards and secondary standards. A PRIMARY STANDARD is a reagent for which we can dispense an accurately known amount of analyte. For example, a 0.1250-g sample of $K_2Cr_2O_7$ contains 4.249×10^{-4} moles of $K_2Cr_2O_7$. If we place this sample in a 250-mL volumetric flask and dilute to volume, the concentration of the resulting solution is 1.700×10^{-3} M. A primary standard must have a known stoichiometry, a known purity (or assay), and it must be stable during long-term storage. Because of the difficulty in establishing the degree of hydration, even after drying, a hydrated reagent usually is not a primary standard.

Reagents that do not meet these criteria are SECONDARY STANDARDS. The concentration of a secondary standard must be determined relative to a primary standard. Lists of acceptable primary standards are available.[2] Appendix 8 provides examples of some common primary standards.

The base NaOH is an example of a secondary standard. Commercially available NaOH contains impurities of NaCl, Na_2CO_3, and Na_2SO_4, and readily absorbs H_2O from the atmosphere. To determine the concentration of NaOH in a solution, it is titrated against a primary standard weak acid, such as potassium hydrogen phthalate, $KHC_8H_4O_4$.

5A.2 Other Reagents

Preparing a standard often requires additional reagents that are not primary standards or secondary standards. Preparing a standard solution, for example, requires a suitable solvent, and additional reagents may be need to adjust the standard's matrix. These solvents and reagents are potential sources of additional analyte, which, if not accounted for, produce a determinate error in the standardization. If available, REAGENT GRADE chemicals conforming to standards set by the American Chemical Society should be used.[3] The label on the bottle of a reagent grade chemical (Figure 5.1) lists either the limits for specific impurities, or provides an assay for the impurities. We can improve the quality of a reagent grade chemical by purifying it, or by conducting a more accurate assay. As discussed later in the chapter, we

2 (a) Smith, B. W.; Parsons, M. L. *J. Chem. Educ.* **1973**, *50*, 679–681; (b) Moody, J. R.; Greenburg, P. R.; Pratt, K. W.; Rains, T. C. *Anal. Chem.* **1988**, *60*, 1203A–1218A.

3 Committee on Analytical Reagents, *Reagent Chemicals*, 8th ed., American Chemical Society: Washington, D. C., 1993.

can correct for contributions to S_{total} from reagents used in an analysis by including an appropriate blank determination in the analytical procedure.

5A.3 Preparing Standard Solutions

It is often necessary to prepare a series of standards, each with a different concentration of analyte. We can prepare these standards in two ways. If the range of concentrations is limited to one or two orders of magnitude, then each solution is best prepared by transferring a known mass or volume of the pure standard to a volumetric flask and diluting to volume.

When working with larger ranges of concentration, particularly those extending over more than three orders of magnitude, standards are best prepared by a SERIAL DILUTION from a single stock solution. In a serial dilution we prepare the most concentrated standard and then dilute a portion of it to prepare the next most concentrated standard. Next, we dilute a portion of the second standard to prepare a third standard, continuing this process until all we have prepared all of our standards. Serial dilutions must be prepared with extra care because an error in preparing one standard is passed on to all succeeding standards.

Figure 5.1 Examples of typical packaging labels for reagent grade chemicals. Label (a) provides the manufacturer's assay for the reagent, NaBr. Note that potassium is flagged with an asterisk (*) because its assay exceeds the limits established by the American Chemical Society (ACS). Label (b) does not provide an assay for impurities, but indicates that the reagent meets ACS specifications. An assay for the reagent, $NaHCO_3$ is provided.

5B Calibrating the Signal (S_{total})

The accuracy of our determination of k_A and S_{reag} depends on how accurately we can measure the signal, S_{total}. We measure signals using equipment, such as glassware and balances, and instrumentation, such as spectrophotometers and pH meters. To minimize determinate errors affecting the signal, we first calibrate our equipment and instrumentation. We accomplish the calibration by measuring S_{total} for a standard with a known response of S_{std}, adjusting S_{total} until

$$S_{total} = S_{std}$$

Here are two examples of how we calibrate signals. Other examples are provided in later chapters focusing on specific analytical methods.

When the signal is a measurement of mass, we determine S_{total} using an analytical balance. To calibrate the balance's signal we use a reference weight that meets standards established by a governing agency, such as the National Institute for Standards and Technology or the American Society for Testing and Materials. An electronic balance often includes an internal calibration weight for routine calibrations, as well as programs for calibrating with external weights. In either case, the balance automatically adjusts S_{total} to match S_{std}.

See Section 2D.1 to review how an electronic balance works. Calibrating a balance is important, but it does not eliminate all sources of determinate error in measuring mass. See Appendix 9 for a discussion of correcting for the buoyancy of air.

We also must calibrate our instruments. For example, we can evaluate a spectrophotometer's accuracy by measuring the absorbance of a carefully prepared solution of 60.06 mg/L $K_2Cr_2O_7$ in 0.0050 M H_2SO_4, using 0.0050 M H_2SO_4 as a reagent blank.[4] An absorbance of 0.640 ± 0.010 absorbance units at a wavelength of 350.0 nm indicates that the spectrometer's signal is properly calibrated. Be sure to read and carefully follow the calibration instructions provided with any instrument you use.

5C Determining the Sensitivity (k_A)

To standardize an analytical method we also must determine the value of k_A in equation 5.1 or equation 5.2.

$$S_{total} = k_A n_A + S_{reag} \qquad 5.1$$

$$S_{total} = k_A C_A + S_{reag} \qquad 5.2$$

In principle, it should be possible to derive the value of k_A for any analytical method by considering the chemical and physical processes generating the signal. Unfortunately, such calculations are not feasible when we lack a sufficiently developed theoretical model of the physical processes, or are not useful because of nonideal chemical behavior. In such situations we must determine the value of k_A by analyzing one or more standard solutions, each containing a known amount of analyte. In this section we consider

4 Ebel, S. *Fresenius J. Anal. Chem.* **1992**, *342*, 769.

several approaches for determining the value of k_A. For simplicity we will assume that S_{reag} has been accounted for by a proper reagent blank, allowing us to replace S_{total} in equation 5.1 and equation 5.2 with the analyte's signal, S_A.

$$S_A = k_A n_A \tag{5.3}$$

$$S_A = k_A C_A \tag{5.4}$$

Equation 5.3 and equation 5.4 are essentially identical, differing only in whether we choose to express the amount of analyte in moles or as a concentration. For the remainder of this chapter we will limit our treatment to equation 5.4. You can extend this treatment to equation 5.3 by replacing C_A with n_A.

5C.1 Single-Point versus Multiple-Point Standardizations

The simplest way to determine the value of k_A in equation 5.4 is by a SINGLE-POINT STANDARDIZATION in which we measure the signal for a standard, S_{std}, containing a known concentration of analyte, C_{std}. Substituting these values into equation 5.4

$$k_A = \frac{S_{std}}{C_{std}} \tag{5.5}$$

gives the value for k_A. Having determined the value for k_A, we can calculate the concentration of analyte in any sample by measuring its signal, S_{samp}, and calculating C_A using equation 5.6.

$$C_A = \frac{S_{samp}}{k_A} \tag{5.6}$$

A single-point standardization is the least desirable method for standardizing a method. There are at least two reasons for this. First, any error in our determination of k_A carries over into our calculation of C_A. Second, our experimental value for k_A is for a single concentration of analyte. Extending this value of k_A to other concentrations of analyte requires us to assume a linear relationship between the signal and the analyte's concentration, an assumption that often is not true.[5] Figure 5.2 shows how assuming a constant value of k_A may lead to a determinate error in the analyte's concentration. Despite these limitations, single-point standardizations find routine use when the expected range for the analyte's concentrations is small. Under these conditions it is often safe to assume that k_A is constant (although you should verify this assumption experimentally). This is the case, for example, in clinical labs where many automated analyzers use only a single standard.

The preferred approach to standardizing a method is to prepare a series of standards, each containing the analyte at a different concentration. Standards are chosen such that they bracket the expected range for the analyte's concentration. A MULTIPLE-POINT STANDARDIZATION should include at least three standards, although more are preferable. A plot of S_{std} versus

Linear regression, which also is known as the method of least squares, is one such algorithm. Its use is covered in Section 5D.

5 Cardone, M. J.; Palmero, P. J.; Sybrandt, L. B. *Anal. Chem.* **1980**, *52*, 1187–1191.

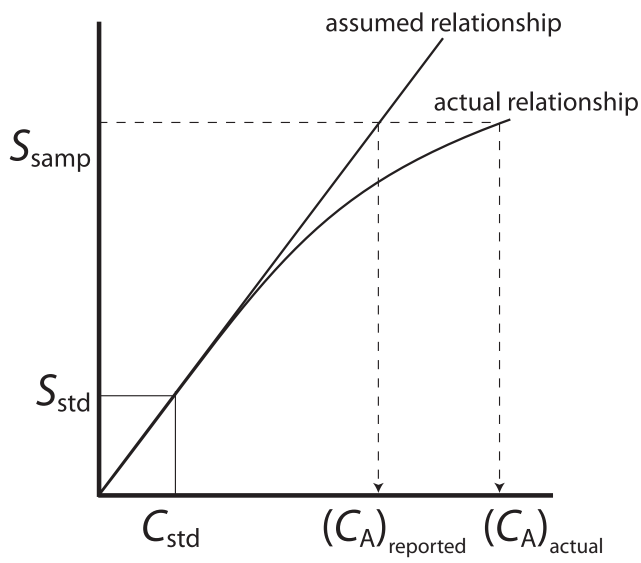


Figure 5.2 Example showing how a single-point standardization leads to a determinate error in an analyte's reported concentration if we incorrectly assume that the value of k_A is constant.

C_{std} is known as a calibration curve. The exact standardization, or calibration relationship is determined by an appropriate curve-fitting algorithm.

There are at least two advantages to a multiple-point standardization. First, although a determinate error in one standard introduces a determinate error into the analysis, its effect is minimized by the remaining standards. Second, by measuring the signal for several concentrations of analyte we no longer must assume that the value of k_A is independent of the analyte's concentration. Constructing a calibration curve similar to the "actual relationship" in Figure 5.2, is possible.

Appending the adjective "external" to the noun "standard" might strike you as odd at this point, as it seems reasonable to assume that standards and samples must be analyzed separately. As you will soon learn, however, we can add standards to our samples and analyze them simultaneously.

5C.2 External Standards

The most common method of standardization uses one or more EXTERNAL STANDARDS, each containing a known concentration of analyte. We call them "external" because we prepare and analyze the standards separate from the samples.

Single External Standard

A quantitative determination using a single external standard was described at the beginning of this section, with k_A given by equation 5.5. After determining the value of k_A, the concentration of analyte, C_A, is calculated using equation 5.6.

Example 5.1

A spectrophotometric method for the quantitative analysis of Pb^{2+} in blood yields an S_{std} of 0.474 for a single standard whose concentration of lead is 1.75 ppb What is the concentration of Pb^{2+} in a sample of blood for which S_{samp} is 0.361?

Solution

Equation 5.5 allows us to calculate the value of k_A for this method using the data for the standard.

$$k_A = \frac{S_{std}}{C_{std}} = \frac{0.474}{1.75 \text{ ppb}} = 0.2709 \text{ ppb}^{-1}$$

Having determined the value of k_A, the concentration of Pb^{2+} in the sample of blood is calculated using equation 5.6.

$$C_A = \frac{S_{samp}}{k_A} = \frac{0.361}{0.2709 \text{ ppb}^{-1}} = 1.33 \text{ ppb}$$

Multiple External Standards

Figure 5.3 shows a typical multiple-point external standardization. The volumetric flask on the left is a reagent blank and the remaining volumetric flasks contain increasing concentrations of Cu^{2+}. Shown below the volumetric flasks is the resulting calibration curve. Because this is the most common method of standardization the resulting relationship is called a NORMAL CALIBRATION CURVE.

When a calibration curve is a straight-line, as it is in Figure 5.3, the slope of the line gives the value of k_A. This is the most desirable situation since the method's sensitivity remains constant throughout the analyte's concentration range. When the calibration curve is not a straight-line, the

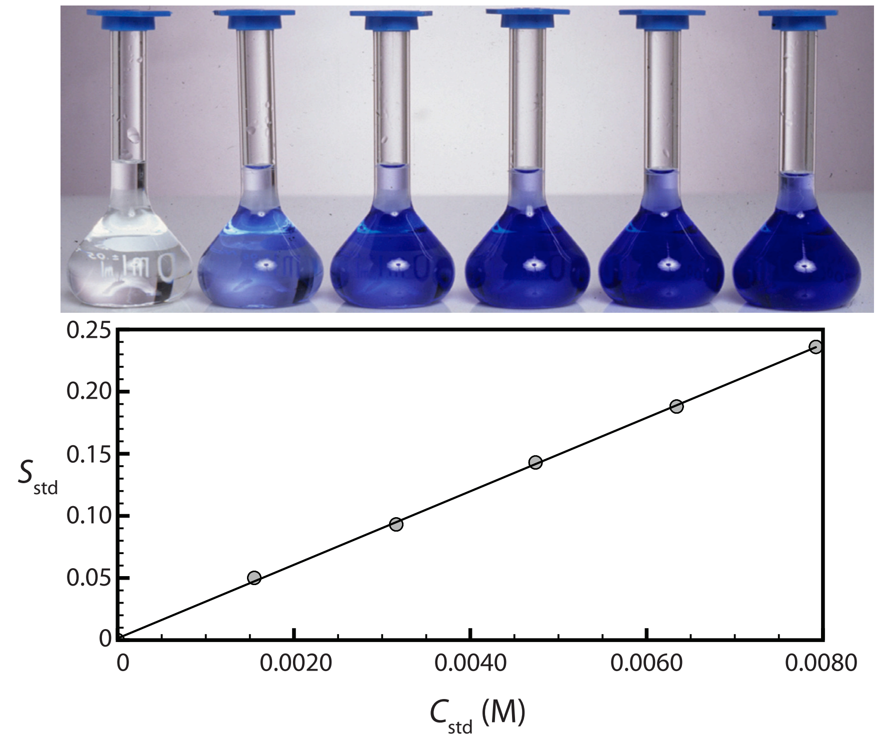


Figure 5.3 Shown at the top is a reagent blank (far left) and a set of five external standards for Cu^{2+} with concentrations increasing from left to right. Shown below the external standards is the resulting normal calibration curve. The absorbance of each standard, S_{std}, is shown by the filled circles.

method's sensitivity is a function of the analyte's concentration. In Figure 5.2, for example, the value of k_A is greatest when the analyte's concentration is small and decreases continuously for higher concentrations of analyte. The value of k_A at any point along the calibration curve in Figure 5.2 is given by the slope at that point. In either case, the calibration curve provides a means for relating S_{samp} to the analyte's concentration.

Example 5.2

A second spectrophotometric method for the quantitative analysis of Pb^{2+} in blood has a normal calibration curve for which

$$S_{std} = (0.296 \text{ ppb}^{-1}) \times C_{std} + 0.003$$

What is the concentration of Pb^{2+} in a sample of blood if S_{samp} is 0.397?

Solution

To determine the concentration of Pb^{2+} in the sample of blood we replace S_{std} in the calibration equation with S_{samp} and solve for C_A.

$$C_A = \frac{S_{samp} - 0.003}{0.296 \text{ ppb}^{-1}} = \frac{0.397 - 0.003}{0.296 \text{ ppb}^{-1}} = 1.33 \text{ ppb}$$

It is worth noting that the calibration equation in this problem includes an extra term that does not appear in equation 5.6. Ideally we expect the calibration curve to have a signal of zero when C_A is zero. This is the purpose of using a reagent blank to correct the measured signal. The extra term of +0.003 in our calibration equation results from the uncertainty in measuring the signal for the reagent blank and the standards.

Practice Exercise 5.1

Figure 5.3 shows a normal calibration curve for the quantitative analysis of Cu^{2+}. The equation for the calibration curve is

$$S_{std} = 29.59 \text{ M}^{-1} \times C_{std} + 0.0015$$

What is the concentration of Cu^{2+} in a sample whose absorbance, S_{samp}, is 0.114? Compare your answer to a one-point standardization where a standard of 3.16×10^{-3} M Cu^{2+} gives a signal of 0.0931.

Click here to review your answer to this exercise.

The one-point standardization in this exercise uses data from the third volumetric flask in Figure 5.3.

An external standardization allows us to analyze a series of samples using a single calibration curve. This is an important advantage when we have many samples to analyze. Not surprisingly, many of the most common quantitative analytical methods use an external standardization.

There is a serious limitation, however, to an external standardization. When we determine the value of k_A using equation 5.5, the analyte is pres-

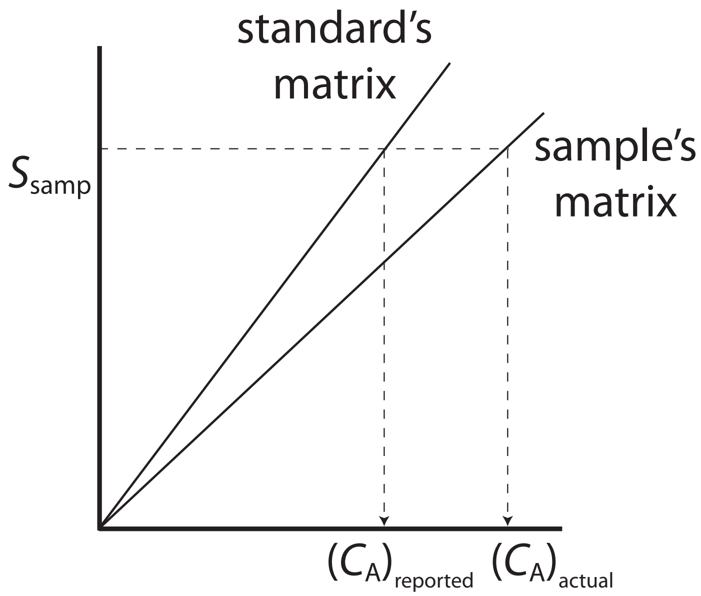


Figure 5.4 Calibration curves for an analyte in the standard's matrix and in the sample's matrix. If the matrix affects the value of k_A, as is the case here, then we introduce a determinate error into our analysis if we use a normal calibration curve.

ent in the external standard's matrix, which usually is a much simpler matrix than that of our samples. When using an external standardization we assume that the matrix does not affect the value of k_A. If this is not true, then we introduce a proportional determinate error into our analysis. This is not the case in Figure 5.4, for instance, where we show calibration curves for the analyte in the sample's matrix and in the standard's matrix. In this example, a calibration curve using external standards results in a negative determinate error. If we expect that matrix effects are important, then we try to match the standard's matrix to that of the sample. This is known as MATRIX MATCHING. If we are unsure of the sample's matrix, then we must show that matrix effects are negligible, or use an alternative method of standardization. Both approaches are discussed in the following section.

The matrix for the external standards in Figure 5.3, for example, is dilute ammonia, which is added because the $Cu(NH_3)_4^{2+}$ complex absorbs more strongly than Cu^{2+}. If we fail to add the same amount of ammonia to our samples, then we will introduce a proportional determinate error into our analysis.

5C.3 Standard Additions

We can avoid the complication of matching the matrix of the standards to the matrix of the sample by conducting the standardization in the sample. This is known as the METHOD OF STANDARD ADDITIONS.

Single Standard Addition

The simplest version of a standard addition is shown in Figure 5.5. First we add a portion of the sample, V_o, to a volumetric flask, dilute it to volume, V_f, and measure its signal, S_{samp}. Next, we add a second identical portion of sample to an equivalent volumetric flask along with a spike, V_{std}, of an external standard whose concentration is C_{std}. After diluting the spiked sample to the same final volume, we measure its signal, S_{spike}. The following two equations relate S_{samp} and S_{spike} to the concentration of analyte, C_A, in the original sample.

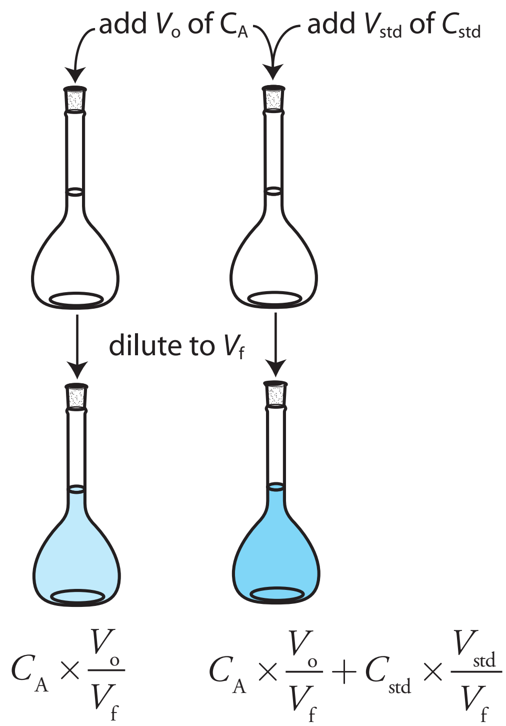


Figure 5.5 Illustration showing the method of standard additions. The volumetric flask on the left contains a portion of the sample, V_o, and the volumetric flask on the right contains an identical portion of the sample and a spike, V_{std}, of a standard solution of the analyte. Both flasks are diluted to the same final volume, V_f. The concentration of analyte in each flask is shown at the bottom of the figure where C_A is the analyte's concentration in the original sample and C_{std} is the concentration of analyte in the external standard.

The ratios V_o/V_f and V_{std}/V_f account for the dilution of the sample and the standard, respectively.

$$S_{samp} = k_A C_A \frac{V_o}{V_f} \tag{5.7}$$

$$S_{spike} = k_A \left(C_A \frac{V_o}{V_f} + C_{std} \frac{V_{std}}{V_f} \right) \tag{5.8}$$

As long as V_{std} is small relative to V_o, the effect of the standard's matrix on the sample's matrix is insignificant. Under these conditions the value of k_A is the same in equation 5.7 and equation 5.8. Solving both equations for k_A and equating gives

$$\frac{S_{samp}}{C_A \frac{V_o}{V_f}} = \frac{S_{spike}}{C_A \frac{V_o}{V_f} + C_{std} \frac{V_{std}}{V_f}} \tag{5.9}$$

which we can solve for the concentration of analyte, C_A, in the original sample.

Example 5.3

A third spectrophotometric method for the quantitative analysis of Pb^{2+} in blood yields an S_{samp} of 0.193 when a 1.00 mL sample of blood is diluted to 5.00 mL. A second 1.00 mL sample of blood is spiked with 1.00 μL of a 1560-ppb Pb^{2+} external standard and diluted to 5.00 mL, yielding an

S_{spike} of 0.419. What is the concentration of Pb^{2+} in the original sample of blood?

Solution

We begin by making appropriate substitutions into equation 5.9 and solving for C_A. Note that all volumes must be in the same units; thus, we first covert V_{std} from 1.00 µL to 1.00×10^{-3} mL.

$$\frac{0.193}{C_A \dfrac{1.00 \text{ mL}}{5.00 \text{ mL}}} = \frac{0.419}{C_A \dfrac{1.00 \text{ mL}}{5.00 \text{ mL}} + 1560 \text{ ppb} \dfrac{1.00 \times 10^{-3} \text{ mL}}{5.00 \text{ mL}}}$$

$$\frac{0.193}{0.200 C_A} = \frac{0.419}{0.200 C_A + 0.3120 \text{ ppb}}$$

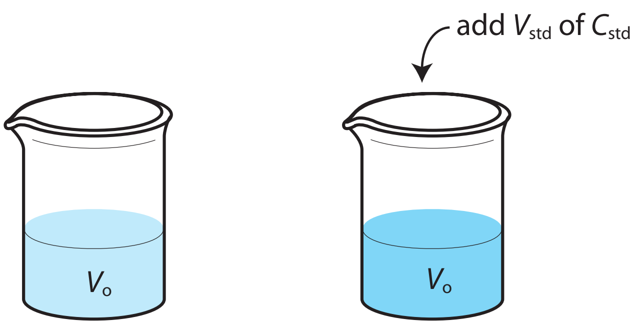
$$0.0386 C_A + 0.0602 \text{ ppb} = 0.0838 C_A$$

$$0.0452 C_A = 0.0602 \text{ ppb}$$

$$C_A = 1.33 \text{ ppb}$$

The concentration of Pb^{2+} in the original sample of blood is 1.33 ppb.

It also is possible to make a standard addition directly to the sample, measuring the signal both before and after the spike (Figure 5.6). In this case the final volume after the standard addition is $V_o + V_{std}$ and equation 5.7, equation 5.8, and equation 5.9 become



Concentration of Analyte $\quad C_A \quad\quad C_A \dfrac{V_o}{V_o + V_{std}} + C_{std} \dfrac{V_{std}}{V_o + V_{std}}$

Figure 5.6 Illustration showing an alternative form of the method of standard additions. In this case we add a spike of the external standard directly to the sample without any further adjust in the volume.

$$S_{samp} = k_A C_A$$

$$S_{spike} = k_A \left(C_A \frac{V_o}{V_o + V_{std}} + C_{std} \frac{V_{std}}{V_o + V_{std}} \right) \tag{5.10}$$

$$\frac{S_{samp}}{C_A} = \frac{S_{spike}}{C_A \frac{V_o}{V_o + V_{std}} + C_{std} \frac{V_{std}}{V_o + V_{std}}} \tag{5.11}$$

Example 5.4

A fourth spectrophotometric method for the quantitative analysis of Pb^{2+} in blood yields an S_{samp} of 0.712 for a 5.00 mL sample of blood. After spiking the blood sample with 5.00 μL of a 1560-ppb Pb^{2+} external standard, an S_{spike} of 1.546 is measured. What is the concentration of Pb^{2+} in the original sample of blood.

Solution

To determine the concentration of Pb^{2+} in the original sample of blood, we make appropriate substitutions into equation 5.11 and solve for C_A.

$$V_o + V_{std} = 5.00 \text{ mL} + 5.00 \times 10^{-3} \text{ mL} = 5.005 \text{ mL}$$

$$\frac{0.712}{C_A} = \frac{1.546}{C_A \frac{5.00 \text{ mL}}{5.005 \text{ mL}} + 1560 \text{ ppb} \frac{5.00 \times 10^{-3} \text{ mL}}{5.005 \text{ mL}}}$$

$$\frac{0.712}{C_A} = \frac{1.546}{0.9990 C_A + 1.558 \text{ ppb}}$$

$$0.7113 C_A + 1.109 \text{ ppb} = 1.546 C_A$$

$$C_A = 1.33 \text{ ppb}$$

The concentration of Pb^{2+} in the original sample of blood is 1.33 ppb.

Multiple Standard Additions

We can adapt the single-point standard addition into a multiple-point standard addition by preparing a series of samples containing increasing amounts of the external standard. Figure 5.7 shows two ways to plot a standard addition calibration curve based on equation 5.8. In Figure 5.7a we plot S_{spike} against the volume of the spikes, V_{std}. If k_A is constant, then the calibration curve is a straight-line. It is easy to show that the x-intercept is equivalent to $-C_A V_o / C_{std}$.

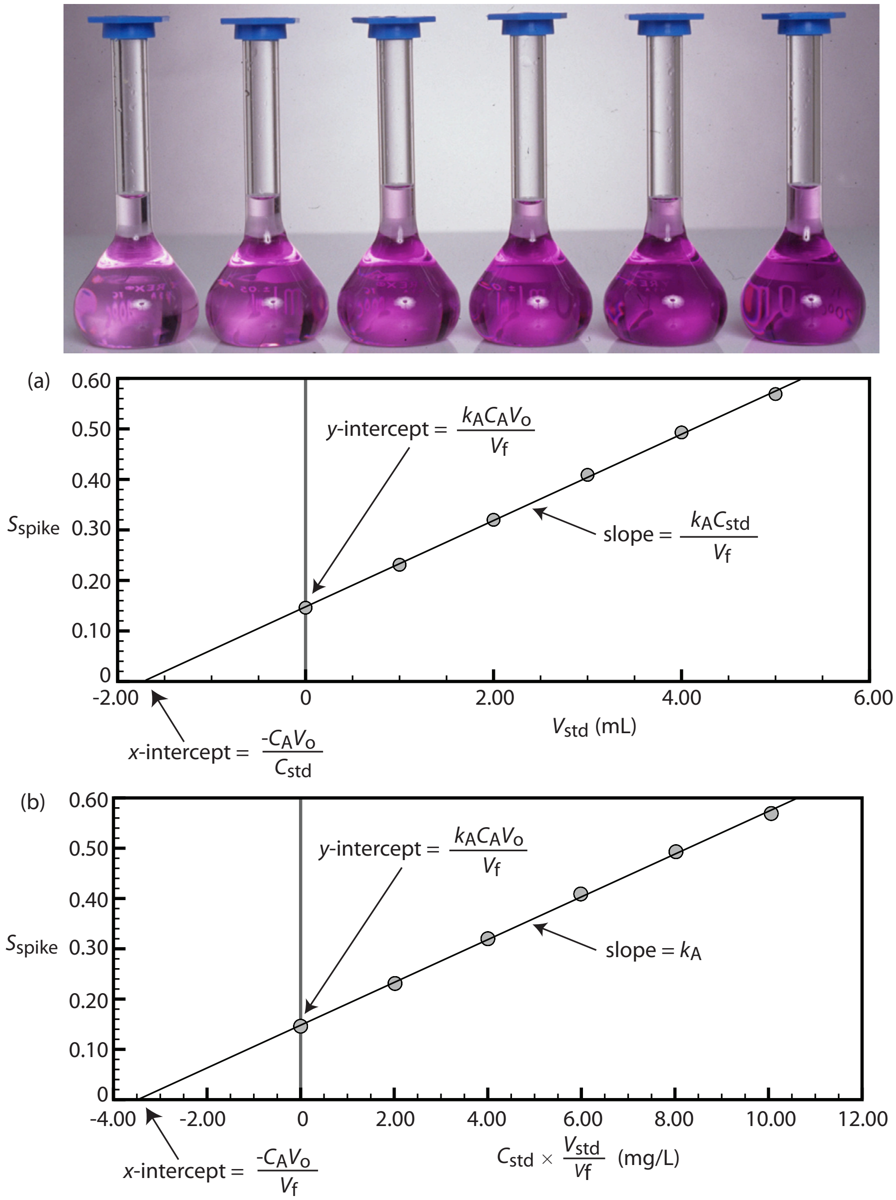


Figure 5.7 Shown at the top is a set of six standard additions for the determination of Mn^{2+}. The flask on the left is a 25.00 mL sample diluted to 50.00 mL. The remaining flasks contain 25.00 mL of sample and, from left to right, 1.00, 2.00, 3.00, 4.00, and 5.00 mL of an external standard of 100.6 mg/L Mn^{2+}. Shown below are two ways to plot the standard additions calibration curve. The absorbance for each standard addition, S_{spike}, is shown by the filled circles.

Example 5.5

Beginning with equation 5.8 show that the equations in Figure 5.7a for the slope, the *y*-intercept, and the *x*-intercept are correct.

Solution

We begin by rewriting equation 5.8 as

$$S_{spike} = \frac{k_A C_A V_o}{V_f} + \frac{k_A C_{std}}{V_f} \times V_{std}$$

which is in the form of the equation for a straight-line

$$Y = y\text{-intercept} + \text{slope} \times X$$

where Y is S_{spike} and X is V_{std}. The slope of the line, therefore, is $k_A C_{std}/V_f$ and the y-intercept is $k_A C_A V_o/V_f$. The x-intercept is the value of X when Y is zero, or

$$0 = \frac{k_A C_A V_o}{V_f} + \frac{k_A C_{std}}{V_f} \times x\text{-intercept}$$

$$x\text{-intercept} = -\frac{k_A C_A V_o \Big/ V_f}{k_A C_{std} \Big/ V_f} = -\frac{C_A V_o}{C_{std}}$$

Practice Exercise 5.2

Beginning with equation 5.8 show that the equations in Figure 5.7b for the slope, the y-intercept, and the x-intercept are correct.

Click here to review your answer to this exercise.

Because we know the volume of the original sample, V_o, and the concentration of the external standard, C_{std}, we can calculate the analyte's concentrations from the x-intercept of a multiple-point standard additions.

Example 5.6

A fifth spectrophotometric method for the quantitative analysis of Pb^{2+} in blood uses a multiple-point standard addition based on equation 5.8. The original blood sample has a volume of 1.00 mL and the standard used for spiking the sample has a concentration of 1560 ppb Pb^{2+}. All samples were diluted to 5.00 mL before measuring the signal. A calibration curve of S_{spike} versus V_{std} has the following equation

$$S_{spike} = 0.266 + 312\ \text{mL}^{-1} \times V_{std}$$

What is the concentration of Pb^{2+} in the original sample of blood.

Solution

To find the x-intercept we set S_{spike} equal to zero.

$$0 = 0.266 + 312\ \text{mL}^{-1} \times V_{std}$$

Solving for V_{std}, we obtain a value of -8.526×10^{-4} mL for the x-intercept. Substituting the x-interecpt's value into the equation from Figure 5.7a

$$-8.526 \times 10^{-4}\ \text{mL} = -\frac{C_A V_o}{C_{std}} = -\frac{C_A \times 1.00\ \text{mL}}{1560\ \text{ppb}}$$

and solving for C_A gives the concentration of Pb^{2+} in the blood sample as 1.33 ppb.

Practice Exercise 5.3

Figure 5.7 shows a standard additions calibration curve for the quantitative analysis of Mn^{2+}. Each solution contains 25.00 mL of the original sample and either 0, 1.00, 2.00, 3.00, 4.00, or 5.00 mL of a 100.6 mg/L external standard of Mn^{2+}. All standard addition samples were diluted to 50.00 mL before reading the absorbance. The equation for the calibration curve in Figure 5.7a is

$$S_{std} = 0.0854 \times V_{std} + 0.1478$$

What is the concentration of Mn^{2+} in this sample? Compare your answer to the data in Figure 5.7b, for which the calibration curve is

$$S_{std} = 0.0425 \times C_{std}(V_{std}/V_f) + 0.1478$$

Click here to review your answer to this exercise.

Since we construct a standard additions calibration curve in the sample, we can not use the calibration equation for other samples. Each sample, therefore, requires its own standard additions calibration curve. This is a serious drawback if you have many samples. For example, suppose you need to analyze 10 samples using a three-point calibration curve. For a normal calibration curve you need to analyze only 13 solutions (three standards and ten samples). If you use the method of standard additions, however, you must analyze 30 solutions (each of the ten samples must be analyzed three times, once before spiking and after each of two spikes).

Using a Standard Addition to Identify Matrix Effects

We can use the method of standard additions to validate an external standardization when matrix matching is not feasible. First, we prepare a normal calibration curve of S_{std} versus C_{std} and determine the value of k_A from its slope. Next, we prepare a standard additions calibration curve using equation 5.8, plotting the data as shown in Figure 5.7b. The slope of this standard additions calibration curve provides an independent determination of k_A. If there is no significant difference between the two values of k_A, then we can ignore the difference between the sample's matrix and that of the external standards. When the values of k_A are significantly different, then using a normal calibration curve introduces a proportional determinate error.

5C.4 Internal Standards

To successfully use an external standardization or the method of standard additions, we must be able to treat identically all samples and standards. When this is not possible, the accuracy and precision of our standardization may suffer. For example, if our analyte is in a volatile solvent, then its concentration increases when we lose solvent to evaporation. Suppose we have a sample and a standard with identical concentrations of analyte and identical signals. If both experience the same proportional loss of solvent then their respective concentrations of analyte and signals continue to be identical. In effect, we can ignore evaporation if the samples and standards experience an equivalent loss of solvent. If an identical standard and sample lose different amounts of solvent, however, then their respective concentrations and signals will no longer be equal. In this case a simple external standardization or standard addition is not possible.

We can still complete a standardization if we reference the analyte's signal to a signal from another species that we add to all samples and standards. The species, which we call an INTERNAL STANDARD, must be different than the analyte.

Because the analyte and the internal standard in any sample or standard receive the same treatment, the ratio of their signals is unaffected by any lack of reproducibility in the procedure. If a solution contains an analyte of concentration C_A, and an internal standard of concentration, C_{IS}, then the signals due to the analyte, S_A, and the internal standard, S_{IS}, are

$$S_A = k_A C_A$$

$$S_{IS} = k_{IS} C_{IS}$$

where k_A and k_{IS} are the sensitivities for the analyte and internal standard. Taking the ratio of the two signals gives the fundamental equation for an internal standardization.

$$\frac{S_A}{S_{IS}} = \frac{k_A C_A}{k_{IS} C_{IS}} = K \times \frac{C_A}{C_{IS}} \qquad 5.12$$

Because K is a ratio of the analyte's sensitivity and the internal standard's sensitivity, it is not necessary to independently determine values for either k_A or k_{IS}.

Single Internal Standard

In a single-point internal standardization, we prepare a single standard containing the analyte and the internal standard, and use it to determine the value of K in equation 5.12.

$$K = \left(\frac{C_{IS}}{C_A}\right)_{std} \times \left(\frac{S_A}{S_{IS}}\right)_{std} \qquad 5.13$$

Having standardized the method, the analyte's concentration is given by

$$C_A = \frac{C_{IS}}{K} \times \left(\frac{S_A}{S_{IS}}\right)_{samp}$$

Example 5.7

A sixth spectrophotometric method for the quantitative analysis of Pb^{2+} in blood uses Cu^{2+} as an internal standard. A standard containing 1.75 ppb Pb^{2+} and 2.25 ppb Cu^{2+} yields a ratio of $(S_A/S_{IS})_{std}$ of 2.37. A sample of blood is spiked with the same concentration of Cu^{2+}, giving a signal ratio, $(S_A/S_{IS})_{samp}$, of 1.80. Determine the concentration of Pb^{2+} in the sample of blood.

Solution

Equation 5.13 allows us to calculate the value of K using the data for the standard

$$K = \left(\frac{C_{IS}}{C_A}\right)_{std} \times \left(\frac{S_A}{S_{IS}}\right)_{std} = \frac{2.25 \text{ ppb Cu}}{1.75 \text{ ppb Pb}^{2+}} \times 2.37 = 3.05 \frac{\text{ppb Cu}^{2+}}{\text{ppb Pb}^{2+}}$$

The concentration of Pb^{2+}, therefore, is

$$C_A = \frac{C_{IS}}{K} \times \left(\frac{S_A}{S_{IS}}\right)_{samp} = \frac{2.25 \text{ ppb Cu}^{2+}}{3.05 \frac{\text{ppb Cu}^{2+}}{\text{ppb Pb}^{2+}}} \times 1.80 = 1.33 \text{ ppb Cu}^{2+}$$

Multiple Internal Standards

A single-point internal standardization has the same limitations as a single-point normal calibration. To construct an internal standard calibration curve we prepare a series of standards, each containing the same concentration of internal standard and a different concentrations of analyte. Under these conditions a calibration curve of $(S_A/S_{IS})_{std}$ versus C_A is linear with a slope of K/C_{IS}.

Although the usual practice is to prepare the standards so that each contains an identical amount of the internal standard, this is not a requirement.

Example 5.8

A seventh spectrophotometric method for the quantitative analysis of Pb^{2+} in blood gives a linear internal standards calibration curve for which

$$\left(\frac{S_A}{S_{IS}}\right)_{std} = (2.11 \text{ ppb}^{-1}) \times C_A - 0.006$$

What is the ppb Pb^{2+} in a sample of blood if $(S_A/S_{IS})_{samp}$ is 2.80?

Solution

To determine the concentration of Pb^{2+} in the sample of blood we replace $(S_A/S_{IS})_{std}$ in the calibration equation with $(S_A/S_{IS})_{samp}$ and solve for C_A.

$$C_A = \frac{\left(\frac{S_A}{S_{IS}}\right)_{samp} + 0.006}{2.11 \text{ ppb}^{-1}} = \frac{2.80 + 0.006}{2.11 \text{ ppb}^{-1}} = 1.33 \text{ ppb}$$

The concentration of Pb^{2+} in the sample of blood is 1.33 ppb.

In some circumstances it is not possible to prepare the standards so that each contains the same concentration of internal standard. This is the case, for example, when preparing samples by mass instead of volume. We can still prepare a calibration curve, however, by plotting $(S_A/S_{IS})_{std}$ versus C_A/C_{IS}, giving a linear calibration curve with a slope of K.

5D Linear Regression and Calibration Curves

In a single-point external standardization we determine the value of k_A by measuring the signal for a single standard containing a known concentration of analyte. Using this value of k_A and the signal for our sample, we then calculate the concentration of analyte in our sample (see Example 5.1). With only a single determination of k_A, a quantitative analysis using a single-point external standardization is straightforward.

A multiple-point standardization presents a more difficult problem. Consider the data in Table 5.1 for a multiple-point external standardization. What is our best estimate of the relationship between S_{std} and C_{std}? It is tempting to treat this data as five separate single-point standardizations, determining k_A for each standard, and reporting the mean value. Despite it simplicity, this is not an appropriate way to treat a multiple-point standardization.

So why is it inappropriate to calculate an average value for k_A as done in Table 5.1? In a single-point standardization we assume that our reagent blank (the first row in Table 5.1) corrects for all constant sources of determinate error. If this is not the case, then the value of k_A from a single-point standardization has a determinate error. Table 5.2 demonstrates how an

Table 5.1 Data for a Hypothetical Multiple-Point External Standardization

C_{std} (arbitrary units)	S_{std} (arbitrary units)	$k_A = S_{std}/C_{std}$
0.000	0.00	—
0.100	12.36	123.6
0.200	24.83	124.2
0.300	35.91	119.7
0.400	48.79	122.0
0.500	60.42	122.8
	mean value for k_A =	122.5

Table 5.2 Effect of a Constant Determinate Error on the Value of k_A From a Single-Point Standardization

C_{std}	S_{std} (without constant error)	$k_A = S_{std}/C_{std}$ (actual)	$(S_{std})_e$ (with constant error)	$k_A = (S_{std})_e/C_{std}$ (apparent)
1.00	1.00	1.00	1.50	1.50
2.00	2.00	1.00	2.50	1.25
3.00	3.00	1.00	3.50	1.17
4.00	4.00	1.00	4.50	1.13
5.00	5.00	1.00	5.50	1.10
	mean k_A (true) =	1.00	mean k_A (apparent) =	1.23

uncorrected constant error affects our determination of k_A. The first three columns show the concentration of analyte in the standards, C_{std}, the signal without any source of constant error, S_{std}, and the actual value of k_A for five standards. As we expect, the value of k_A is the same for each standard. In the fourth column we add a constant determinate error of +0.50 to the signals, $(S_{std})_e$. The last column contains the corresponding apparent values of k_A. Note that we obtain a different value of k_A for each standard and that all of the apparent k_A values are greater than the true value.

How do we find the best estimate for the relationship between the signal and the concentration of analyte in a multiple-point standardization? Figure 5.8 shows the data in Table 5.1 plotted as a normal calibration curve. Although the data certainly appear to fall along a straight line, the actual calibration curve is not intuitively obvious. The process of mathematically determining the best equation for the calibration curve is called linear regression.

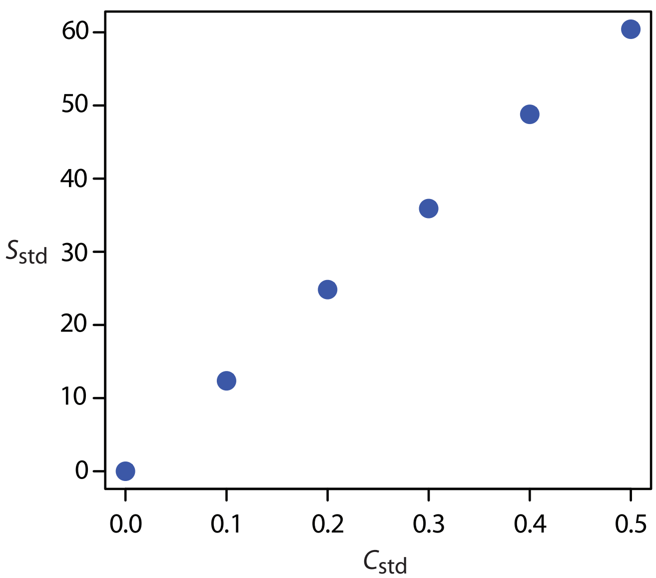


Figure 5.8 Normal calibration curve for the hypothetical multiple-point external standardization in Table 5.1.

5D.1 Linear Regression of Straight Line Calibration Curves

When a calibration curve is a straight-line, we represent it using the following mathematical equation

$$y = \beta_0 + \beta_1 x \tag{5.14}$$

where y is the signal, S_{std}, and x is the analyte's concentration, C_{std}. The constants β_0 and β_1 are, respectively, the calibration curve's expected y-intercept and its expected slope. Because of uncertainty in our measurements, the best we can do is to estimate values for β_0 and β_1, which we represent as b_0 and b_1. The goal of a LINEAR REGRESSION analysis is to determine the best estimates for b_0 and b_1. How we do this depends on the uncertainty in our measurements.

5D.2 Unweighted Linear Regression with Errors in y

The most common approach to completing a linear regression for equation 5.14 makes three assumptions:

(1) that any difference between our experimental data and the calculated regression line is the result of indeterminate errors affecting y,

(2) that indeterminate errors affecting y are normally distributed, and

(3) that the indeterminate errors in y are independent of the value of x.

Because we assume that the indeterminate errors are the same for all standards, each standard contributes equally in estimating the slope and the y-intercept. For this reason the result is considered an UNWEIGHTED LINEAR REGRESSION.

The second assumption is generally true because of the central limit theorem, which we considered in Chapter 4. The validity of the two remaining assumptions is less obvious and you should evaluate them before accepting the results of a linear regression. In particular the first assumption is always suspect since there will certainly be some indeterminate errors affecting the values of x. When preparing a calibration curve, however, it is not unusual for the uncertainty in the signal, S_{std}, to be significantly larger than that for the concentration of analyte in the standards C_{std}. In such circumstances the first assumption is usually reasonable.

How a Linear Regression Works

To understand the logic of an linear regression consider the example shown in Figure 5.9, which shows three data points and two possible straight-lines that might reasonably explain the data. How do we decide how well these straight-lines fits the data, and how do we determine the best straight-line?

Let's focus on the solid line in Figure 5.9. The equation for this line is

$$\hat{y} = b_0 + b_1 x \tag{5.15}$$

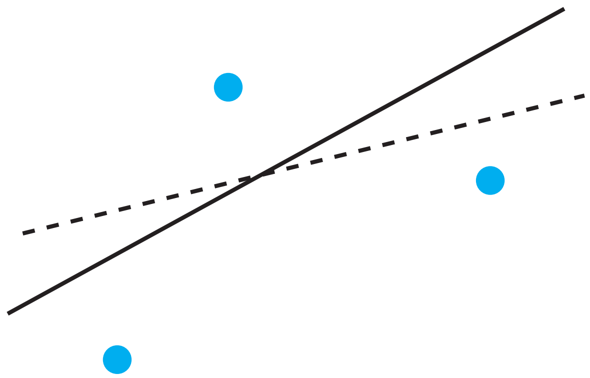


Figure 5.9 Illustration showing three data points and two possible straight-lines that might explain the data. The goal of a linear regression is to find the mathematical model, in this case a straight-line, that best explains the data.

where b_0 and b_1 are our estimates for the y-intercept and the slope, and $\hat{y}$ is our prediction for the experimental value of y for any value of x. Because we assume that all uncertainty is the result of indeterminate errors affecting y, the difference between y and $\hat{y}$ for each data point is the RESIDUAL ERROR, r, in the our mathematical model for a particular value of x.

If you are reading this aloud, you pronounce $\hat{y}$ as y-hat.

$$r_i = (y_i - \hat{y}_i)$$

Figure 5.10 shows the residual errors for the three data points. The smaller the total residual error, R, which we define as

$$R = \sum_i (y_i - \hat{y}_i)^2 \qquad 5.16$$

the better the fit between the straight-line and the data. In a linear regression analysis, we seek values of b_0 and b_1 that give the smallest total residual error.

The reason for squaring the individual residual errors is to prevent positive residual error from canceling out negative residual errors. You have seen this before in the equations for the sample and population standard deviations. You also can see from this equation why a linear regression is sometimes called the method of least squares.

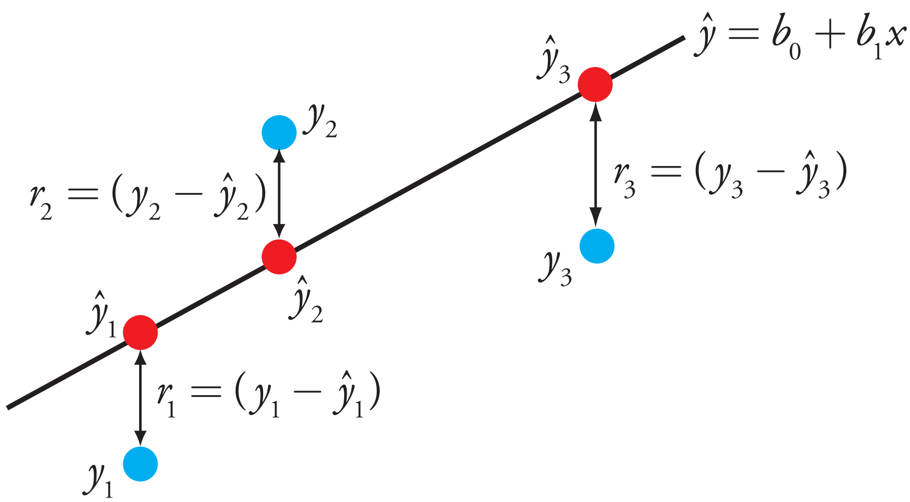


Figure 5.10 Illustration showing the evaluation of a linear regression in which we assume that all uncertainty is the result of indeterminate errors affecting y. The points in blue, y_i, are the original data and the points in red, $\hat{y}_i$, are the predicted values from the regression equation, $\hat{y} = b_0 + b_1 x$. The smaller the total residual error (equation 5.16), the better the fit of the straight-line to the data.

Finding the Slope and y-Intercept

Although we will not formally develop the mathematical equations for a linear regression analysis, you can find the derivations in many standard statistical texts.[6] The resulting equation for the slope, b_1, is

$$b_1 = \frac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \qquad 5.17$$

and the equation for the y-intercept, b_0, is

$$b_0 = \frac{\sum_i y_i - b_1 \sum_i x_i}{n} \qquad 5.18$$

Although equation 5.17 and equation 5.18 appear formidable, it is only necessary to evaluate the following four summations

$$\sum_i x_i \qquad \sum_i y_i \qquad \sum_i x_i y_i \qquad \sum_i x_i^2$$

See Section 5F in this chapter for details on completing a linear regression analysis using Excel and R.

Many calculators, spreadsheets, and other statistical software packages are capable of performing a linear regression analysis based on this model. To save time and to avoid tedious calculations, learn how to use one of these tools. For illustrative purposes the necessary calculations are shown in detail in the following example.

Equations 5.17 and 5.18 are written in terms of the general variables x and y. As you work through this example, remember that x corresponds to C_{std}, and that y corresponds to S_{std}.

Example 5.9

Using the data from Table 5.1, determine the relationship between S_{std} and C_{std} using an unweighted linear regression.

Solution

We begin by setting up a table to help us organize the calculation.

x_i	y_i	$x_i y_i$	x_i^2
0.000	0.00	0.000	0.000
0.100	12.36	1.236	0.010
0.200	24.83	4.966	0.040
0.300	35.91	10.773	0.090
0.400	48.79	19.516	0.160
0.500	60.42	30.210	0.250

Adding the values in each column gives

$$\sum_i x_i = 1.500 \qquad \sum_i y_i = 182.31 \qquad \sum_i x_i y_i = 66.701 \qquad \sum_i x_i^2 = 0.550$$

6 See, for example, Draper, N. R.; Smith, H. Applied Regression Analysis, 3rd ed.; Wiley: New York, 1998.

Substituting these values into equation 5.17 and equation 5.18, we find that the slope and the y-intercept are

$$b_1 = \frac{(6 \times 66.701) - (1.500 \times 182.31)}{(6 \times 0.550) - (1.500)^2} = 120.706 \approx 120.71$$

$$b_0 = \frac{182.31 - (120.706 \times 1.500)}{6} = 0.209 \approx 0.21$$

The relationship between the signal and the analyte, therefore, is

$$S_{\text{std}} = 120.71 \times C_{\text{std}} + 0.21$$

For now we keep two decimal places to match the number of decimal places in the signal. The resulting calibration curve is shown in Figure 5.11.

Uncertainty in the Regression Analysis

As shown in Figure 5.11, because of indeterminate error affecting our signal, the regression line may not pass through the exact center of each data point. The cumulative deviation of our data from the regression line—that is, the total residual error—is proportional to the uncertainty in the regression. We call this uncertainty the STANDARD DEVIATION ABOUT THE REGRESSION, s_r, which is equal to

$$s_r = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n-2}} \qquad 5.19$$

Did you notice the similarity between the standard deviation about the regression (equation 5.19) and the standard deviation for a sample (equation 4.1)?

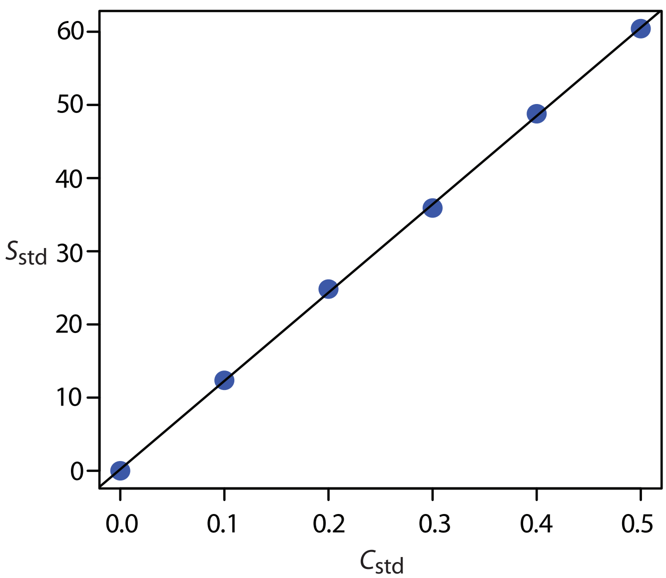


Figure 5.11 Calibration curve for the data in Table 5.1 and Example 5.9.

where y_i is the i_{th} experimental value, and $\hat{y}_i$ is the corresponding value predicted by the regression line in equation 5.15. Note that the denominator of equation 5.19 indicates that our regression analysis has n–2 degrees of freedom—we lose two degree of freedom because we use two parameters, the slope and the y-intercept, to calculate $\hat{y}_i$.

A more useful representation of the uncertainty in our regression is to consider the effect of indeterminate errors on the slope, b_1, and the y-intercept, b_0, which we express as standard deviations.

$$s_{b_1} = \sqrt{\frac{ns_r^2}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}} = \sqrt{\frac{s_r^2}{\sum_i \left(x_i - \bar{x}\right)^2}} \qquad 5.20$$

$$s_{b_0} = \sqrt{\frac{s_r^2 \sum_i x_i^2}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}} = \sqrt{\frac{s_r^2 \sum_i x_i^2}{n\sum_i \left(x_i - \bar{x}\right)^2}} \qquad 5.21$$

We use these standard deviations to establish confidence intervals for the expected slope, β_1, and the expected y-intercept, β_0

$$\beta_1 = b_1 \pm ts_{b_1} \qquad 5.22$$

$$\beta_0 = b_0 \pm ts_{b_0} \qquad 5.23$$

You might contrast this with equation 4.12 for the confidence interval around a sample's mean value.

where we select t for a significance level of α and for n–2 degrees of freedom. Note that equation 5.22 and equation 5.23 do not contain a factor of $(\sqrt{n})^{-1}$ because the confidence interval is based on a single regression line.

Again, many calculators, spreadsheets, and computer software packages provide the standard deviations and confidence intervals for the slope and y-intercept. Example 5.10 illustrates the calculations.

Example 5.10

Calculate the 95% confidence intervals for the slope and y-intercept from Example 5.9.

Solution

As you work through this example, remember that x corresponds to C_{std}, and that y corresponds to S_{std}.

We begin by calculating the standard deviation about the regression. To do this we must calculate the predicted signals, $\hat{y}_i$, using the slope and y-intercept from Example 5.9, and the squares of the residual error, $(y_i - \hat{y}_i)^2$. Using the last standard as an example, we find that the predicted signal is

$$\hat{y}_6 = b_0 + b_1 x_6 = 0.209 + \left(120.706 \times 0.500\right) = 60.562$$

and that the square of the residual error is

$$(y_i - \hat{y}_i)^2 = (60.42 - 60.562)^2 = 0.2016 \approx 0.202$$

The following table displays the results for all six solutions.

x_i	y_i	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$
0.000	0.00	0.209	0.0437
0.100	12.36	12.280	0.0064
0.200	24.83	24.350	0.2304
0.300	35.91	36.421	0.2611
0.400	48.79	48.491	0.0894
0.500	60.42	60.562	0.0202

Adding together the data in the last column gives the numerator of equation 5.19 as 0.6512. The standard deviation about the regression, therefore, is

$$s_r = \sqrt{\frac{0.6512}{6-2}} = 0.4035$$

Next we calculate the standard deviations for the slope and the y-intercept using equation 5.20 and equation 5.21. The values for the summation terms are from in Example 5.9.

$$s_{b_1} = \sqrt{\frac{n s_r^2}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}} = \sqrt{\frac{6 \times (0.4035)^2}{(6 \times 0.550) - (1.550)^2}} = 0.965$$

$$s_{b_0} = \sqrt{\frac{s_r^2 \sum_i x_i^2}{n\sum_i x_i^2 - \left(\sum_i x_i\right)^2}} = \sqrt{\frac{(0.4035)^2 \times 0.550}{(6 \times 0.550) - (1.550)^2}}$$

Finally, the 95% confidence intervals ($\alpha = 0.05$, 4 degrees of freedom) for the slope and y-intercept are

You can find values for t in Appendix 4.

$$\beta_1 = b_1 \pm t s_{b_1} = 120.706 \pm (2.78 \times 0.965) = 120.7 \pm 2.7$$

$$\beta_0 = b_0 \pm t s_{b_0} = 0.209 \pm (2.78 \times 0.292) = 0.2 \pm 0.8$$

The standard deviation about the regression, s_r, suggests that the signal, S_{std}, is precise to one decimal place. For this reason we report the slope and the y-intercept to a single decimal place.

Minimizing Uncertainty in Calibration Curves

To minimize the uncertainty in a calibration curve's slope and y-intercept, you should evenly space your standards over a wide range of analyte concentrations. A close examination of equation 5.20 and equation 5.21 will help you appreciate why this is true. The denominators of both equations include the term $\sum (x_i - \overline{x})^2$. The larger the value of this term—which you accomplish by increasing the range of x around its mean value—the smaller the standard deviations in the slope and the y-intercept. Furthermore, to minimize the uncertainty in the y-intercept, it also helps to decrease the value of the term $\sum x_i$ in equation 5.21, which you accomplish by including standards for lower concentrations of the analyte.

Obtaining the Analyte's Concentration From a Regression Equation

Once we have our regression equation, it is easy to determine the concentration of analyte in a sample. When using a normal calibration curve, for example, we measure the signal for our sample, S_{samp}, and calculate the analyte's concentration, C_A, using the regression equation.

$$C_A = \frac{S_{samp} - b_0}{b_1} \qquad 5.24$$

What is less obvious is how to report a confidence interval for C_A that expresses the uncertainty in our analysis. To calculate a confidence interval we need to know the standard deviation in the analyte's concentration, s_{C_A}, which is given by the following equation

$$s_{C_A} = \frac{s_r}{b_1}\sqrt{\frac{1}{m} + \frac{1}{n} + \frac{\left(\overline{S}_{samp} - \overline{S}_{std}\right)^2}{(b_1)^2 \sum_i \left(C_{std_i} - \overline{C}_{std}\right)^2}} \qquad 5.25$$

Equation 5.25 is written in terms of a calibration experiment. A more general form of the equation, written in terms of x and y, is given here.

$$s_x = \frac{s_r}{b_1}\sqrt{\frac{1}{m} + \frac{1}{n} + \frac{\left(\overline{Y} - \overline{y}\right)^2}{(b_1)^2 \sum_i \left(x_i - \overline{x}\right)^2}}$$

A close examination of equation 5.25 should convince you that the uncertainty in C_A is smallest when the sample's average signal, $\overline{S}_{samp}$, is equal to the average signal for the standards, $\overline{S}_{std}$. When practical, you should plan your calibration curve so that S_{samp} falls in the middle of the calibration curve.

where m is the number of replicate used to establish the sample's average signal ($\overline{S}_{samp}$), n is the number of calibration standards, $\overline{S}_{std}$ is the average signal for the calibration standards, and C_{std_i} and $\overline{C}_{std}$ are the individual and mean concentrations for the calibration standards.[7] Knowing the value of s_{C_A}, the confidence interval for the analyte's concentration is

$$\mu_{C_A} = C_A \pm t s_{C_A}$$

where μ_{C_A} is the expected value of C_A in the absence of determinate errors, and with the value of t based on the desired level of confidence and n–2 degrees of freedom.

7 (a) Miller, J. N. *Analyst* **1991**, *116*, 3–14; (b) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*, Wiley-Interscience: New York, 1986, pp. 126-127; (c) Analytical Methods Committee "Uncertainties in concentrations estimated from calibration experiments," AMC Technical Brief, March 2006 (http://www.rsc.org/images/Brief22_tcm18-51117.pdf)

Example 5.11

Three replicate analyses for a sample containing an unknown concentration of analyte, yield values for S_{samp} of 29.32, 29.16 and 29.51. Using the results from Example 5.9 and Example 5.10, determine the analyte's concentration, C_A, and its 95% confidence interval.

Solution

The average signal, $\bar{S}_{samp}$, is 29.33, which, using equation 5.24 and the slope and the y-intercept from Example 5.9, gives the analyte's concentration as

$$C_A = \frac{\bar{S}_{samp} - b_0}{b_1} = \frac{29.33 - 0.209}{120.706} = 0.241$$

To calculate the standard deviation for the analyte's concentration we must determine the values for $\bar{S}_{std}$ and $\sum\left(C_{std_i} - \bar{C}_{std}\right)^2$. The former is just the average signal for the calibration standards, which, using the data in Table 5.1, is 30.385. Calculating $\sum\left(C_{std_i} - \bar{C}_{std}\right)^2$ looks formidable, but we can simplify its calculation by recognizing that this sum of squares term is the numerator in a standard deviation equation; thus,

$$\sum\left(C_{std_i} - \bar{C}_{std}\right)^2 = \left(s_{C_{std}}\right)^2 \times (n-1)$$

where $s_{C_{std}}$ is the standard deviation for the concentration of analyte in the calibration standards. Using the data in Table 5.1 we find that $s_{C_{std}}$ is 0.1871 and

$$\sum\left(C_{std_i} - \bar{C}_{std}\right)^2 = (0.1871)^2 \times (6-1) = 0.175$$

Substituting known values into equation 5.25 gives

$$s_{C_A} = \frac{0.4035}{120.706}\sqrt{\frac{1}{3} + \frac{1}{6} + \frac{(29.33 - 30.385)^2}{(120.706)^2 \times 0.175}} = 0.0024$$

Finally, the 95% confidence interval for 4 degrees of freedom is

$$\mu_{C_A} = C_A \pm ts_{C_A} = 0.241 \pm (2.78 \times 0.0024) = 0.241 \pm 0.007$$

You can find values for t in Appendix 4.

Figure 5.12 shows the calibration curve with curves showing the 95% confidence interval for C_A.

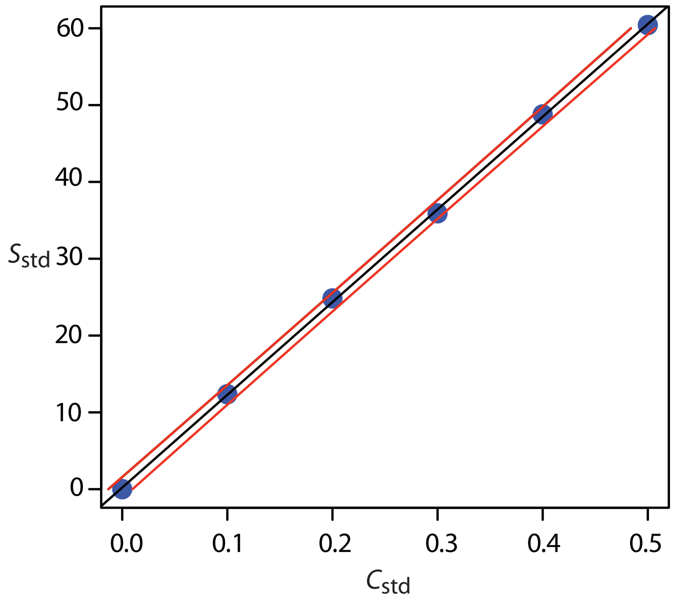


Figure 5.12 Example of a normal calibration curve with a superimposed confidence interval for the analyte's concentration. The points in blue are the original data from Table 5.1. The black line is the normal calibration curve as determined in Example 5.9. The red lines show the 95% confidence interval for C_A assuming a single determination of S_{samp}.

Practice Exercise 5.4

Figure 5.3 shows a normal calibration curve for the quantitative analysis of Cu^{2+}. The data for the calibration curve are shown here.

$[Cu^{2+}]$ (M)	Absorbance
0	0
1.55×10^{-3}	0.050
3.16×10^{-3}	0.093
4.74×10^{-3}	0.143
6.34×10^{-3}	0.188
7.92×10^{-3}	0.236

Complete a linear regression analysis for this calibration data, reporting the calibration equation and the 95% confidence interval for the slope and the *y*-intercept. If three replicate samples give an S_{samp} of 0.114, what is the concentration of analyte in the sample and its 95% confidence interval?

Click here to review your answer to this exercise.

In a standard addition we determine the analyte's concentration by extrapolating the calibration curve to the *x*-intercept. In this case the value of C_A is

$$C_A = x\text{-intercept} = \frac{-b_0}{b_1}$$

and the standard deviation in C_A is

$$s_{C_A} = \frac{s_r}{b_1}\sqrt{\frac{1}{n} + \frac{\left(\overline{S}_{std}\right)^2}{(b_1)^2 \sum_i \left(C_{std_i} - \overline{C}_{std}\right)^2}}$$

where n is the number of standard additions (including the sample with no added standard), and $\overline{S}_{std}$ is the average signal for the n standards. Because we determine the analyte's concentration by extrapolation, rather than by interpolation, s_{C_A} for the method of standard additions generally is larger than for a normal calibration curve.

Evaluating a Linear Regression Model

You should never accept the result of a linear regression analysis without evaluating the validity of the your model. Perhaps the simplest way to evaluate a regression analysis is to examine the residual errors. As we saw earlier, the residual error for a single calibration standard, r_i, is

$$r_i = (y_i - \hat{y}_i)$$

If your regression model is valid, then the residual errors should be randomly distributed about an average residual error of zero, with no apparent trend toward either smaller or larger residual errors (Figure 5.13a). Trends such as those shown in Figure 5.13b and Figure 5.13c provide evidence that at least one of the model's assumptions is incorrect. For example, a trend toward larger residual errors at higher concentrations, as shown in Figure 5.13b, suggests that the indeterminate errors affecting the signal are not independent of the analyte's concentration. In Figure 5.13c, the residual

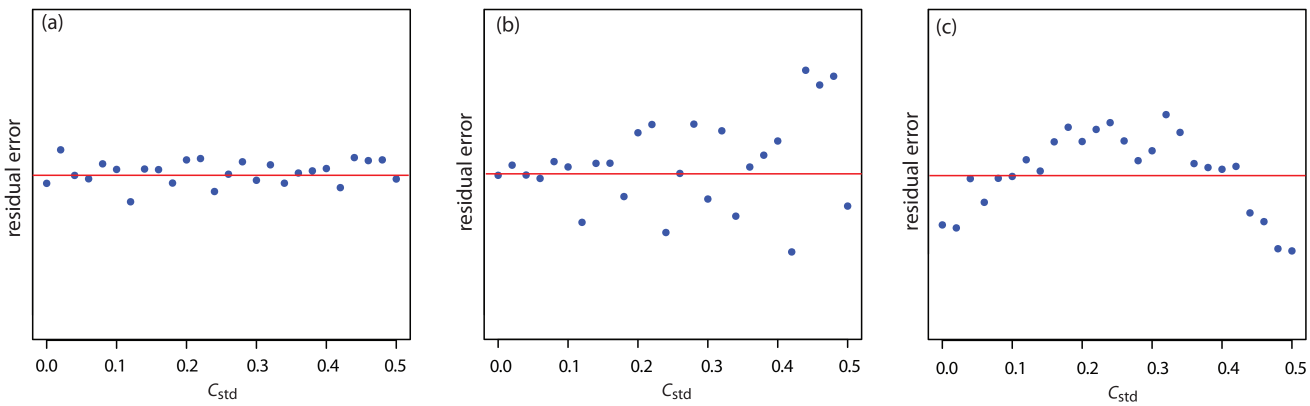


Figure 5.13 Plot of the residual error in the signal, S_{std}, as a function of the concentration of analyte, C_{std} for an unweighted straight-line regression model. The red line shows a residual error of zero. The distribution of the residual error in (a) indicates that the unweighted linear regression model is appropriate. The increase in the residual errors in (b) for higher concentrations of analyte, suggest that a weighted straight-line regression is more appropriate. For (c), the curved pattern to the residuals suggests that a straight-line model is inappropriate; linear regression using a quadratic model might produce a better fit.

Practice Exercise 5.5

Using your results from Practice Exercise 5.4, construct a residual plot and explain its significance.

Click here to review your answer to this exercise.

errors are not random, suggesting that the data can not be modeled with a straight-line relationship. Regression methods for these two cases are discussed in the following sections.

5D.3 Weighted Linear Regression with Errors in y

Our treatment of linear regression to this point assumes that indeterminate errors affecting y are independent of the value of x. If this assumption is false, as is the case for the data in Figure 5.13b, then we must include the variance for each value of y into our determination of the y-intercept, b_o, and the slope, b_1; thus

$$b_0 = \frac{\sum_i w_i y_i - b_1 \sum_i w_i x_i}{n} \qquad 5.26$$

$$b_1 = \frac{n \sum_i w_i x_i y_i - \sum_i w_i x_i \sum_i w_i y_i}{n \sum_i w_i x_i^2 - \left(\sum_i w_i x_i\right)^2} \qquad 5.27$$

where w_i is a weighting factor that accounts for the variance in y_i

$$w_i = \frac{n\left(s_{y_i}\right)^{-2}}{\sum_i \left(s_{y_i}\right)^{-2}} \qquad 5.28$$

and s_{y_i} is the standard deviation for y_i. In a WEIGHTED LINEAR REGRESSION, each xy-pair's contribution to the regression line is inversely proportional to the precision of y_i—that is, the more precise the value of y, the greater its contribution to the regression.

This is the same data used in Example 5.9 with additional information about the standard deviations in the signal.

Example 5.12

Shown here are data for an external standardization in which s_{std} is the standard deviation for three replicate determination of the signal.

C_{std} (arbitrary units)	S_{std} (arbitrary units)	s_{std}
0.000	0.00	0.02
0.100	12.36	0.02
0.200	24.83	0.07

0.300	35.91	0.13
0.400	48.79	0.22
0.500	60.42	0.33

Determine the calibration curve's equation using a weighted linear regression.

As you work through this example, remember that x corresponds to C_{std}, and that y corresponds to S_{std}.

Solution

We begin by setting up a table to aid in calculating the weighting factors.

x_i	y_i	s_{y_i}	$\left(s_{y_i}\right)^{-2}$	w_i
0.000	0.00	0.02	2500.00	2.8339
0.100	12.36	0.02	2500.00	2.8339
0.200	24.83	0.07	204.08	0.2313
0.300	35.91	0.13	59.17	0.0671
0.400	48.79	0.22	20.66	0.0234
0.500	60.42	0.33	9.18	0.0104

As a check on your calculations, the sum of the individual weights must equal the number of calibration standards, n. The sum of the entries in the last column is 6.0000, so all is well.

Adding together the values in the forth column gives

$$\sum_i \left(s_{y_i}\right)^2 = 5293.09$$

which we use to calculate the individual weights in the last column. After calculating the individual weights, we use a second table to aid in calculating the four summation terms in equation 5.26 and equation 5.27.

x_i	y_i	w_i	$w_i x_i$	$w_i y_i$	$w_i x_i^2$	$w_i x_i y_i$
0.000	0.00	2.8339	0.0000	0.0000	0.0000	0.0000
0.100	12.36	2.8339	0.2834	35.0270	0.0283	3.5027
0.200	24.83	0.2313	0.0463	5.7432	0.0093	1.1486
0.300	35.91	0.0671	0.0201	2.4096	0.0060	0.7229
0.400	48.79	0.0234	0.0094	1.1417	0.0037	0.4567
0.500	60.42	0.0104	0.0052	0.6284	0.0026	0.3142

Adding the values in the last four columns gives

$$\sum_i w_i x_i = 0.3644 \quad \sum_i w_i y_i = 44.9499$$

$$\sum_i w_i x_i^2 = 0.0499 \quad \sum_i w_i x_i y_i = 6.1451$$

Substituting these values into the equation 5.26 and equation 5.27 gives the estimated slope and estimated y-intercept as

$$b_1 = \frac{(6 \times 6.1451) - (0.3644 \times 44.9499)}{(6 \times 0.0499) - (0.3644)^2} = 122.985$$

$$b_0 = \frac{44.9499 - (122.985 \times 0.3644)}{6} = 0.0224$$

The calibration equation is

$$S_{std} = 122.98 \times C_{std} + 0.02$$

Figure 5.14 shows the calibration curve for the weighted regression and the calibration curve for the unweighted regression in Example 5.9. Although the two calibration curves are very similar, there are slight differences in the slope and in the y-intercept. Most notably, the y-intercept for the weighted linear regression is closer to the expected value of zero. Because the standard deviation for the signal, S_{std}, is smaller for smaller concentrations of analyte, C_{std}, a weighted linear regression gives more emphasis to these standards, allowing for a better estimate of the y-intercept.

Equations for calculating confidence intervals for the slope, the y-intercept, and the concentration of analyte when using a weighted linear regression are not as easy to define as for an unweighted linear regression.[8] The confidence interval for the analyte's concentration, however, is at its

8 Bonate, P. J. *Anal. Chem.* **1993**, *65*, 1367–1372.

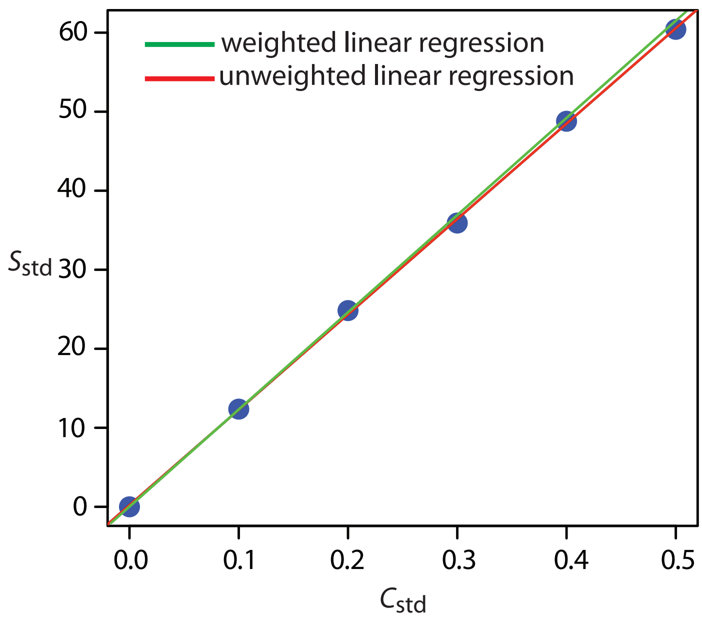


Figure 5.14 A comparison of unweighted and weighted normal calibration curves. See Example 5.9 for details of the unweighted linear regression and Example 5.12 for details of the weighted linear regression.

optimum value when the analyte's signal is near the weighted centroid, y_c, of the calibration curve.

$$y_c = \frac{1}{n}\sum_i w_i x_i$$

5D.4 Weighted Linear Regression with Errors in Both x and y

If we remove our assumption that the indeterminate errors affecting a calibration curve exist only in the signal (y), then we also must factor into the regression model the indeterminate errors affecting the analyte's concentration in the calibration standards (x). The solution for the resulting regression line is computationally more involved than that for either the unweighted or weighted regression lines.[9] Although we will not consider the details in this textbook, you should be aware that neglecting the presence of indeterminate errors in x can bias the results of a linear regression.

See Figure 5.2 for an example of a calibration curve that deviates from a straight-line for higher concentrations of analyte.

5D.5 Curvilinear and Multivariate Regression

A straight-line regression model, despite its apparent complexity, is the simplest functional relationship between two variables. What do we do if our calibration curve is curvilinear—that is, if it is a curved-line instead of a straight-line? One approach is to try transforming the data into a straight-line. Logarithms, exponentials, reciprocals, square roots, and trigonometric functions have been used in this way. A plot of log(y) versus x is a typical example. Such transformations are not without complications. Perhaps the most obvious complication is that data with a uniform variance in y will not maintain that uniform variance after the transformation.

Another approach to developing a linear regression model is to fit a polynomial equation to the data, such as $y = a + bx + cx^2$. You can use linear regression to calculate the parameters a, b, and c, although the equations are different than those for the linear regression of a straight line.[10] If you cannot fit your data using a single polynomial equation, it may be possible to fit separate polynomial equations to short segments of the calibration curve. The result is a single continuous calibration curve known as a spline function.

The regression models in this chapter apply only to functions containing a single independent variable, such as a signal that depends upon the analyte's concentration. In the presence of an interferent, however, the signal may depend on the concentrations of both the analyte and the interferent

It is worth noting that in mathematics, the term "linear" does not mean a straight-line. A linear function may contain many additive terms, but each term can have one and only one adjustable parameter. The function

$$y = ax + bx^2$$

is linear, but the function

$$y = ax^b$$

is nonlinear. This is why you can use linear regression to fit a polynomial equation to your data.

Sometimes it is possible to transform a nonlinear function. For example, taking the log of both sides of the nonlinear function shown above gives a linear function.

$$\log(y) = \log(a) + b\log(x)$$

9 See, for example, Analytical Methods Committee, "Fitting a linear functional relationship to data with error on both variable," AMC Technical Brief, March, 2002 (http://www.rsc.org/images/brief10_tcm18-25920.pdf).

10 For details about curvilinear regression, see (a) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*, Wiley-Interscience: New York, 1986; (b) Deming, S. N.; Morgan, S. L. *Experimental Design: A Chemometric Approach*, Elsevier: Amsterdam, 1987.

$$S = k_A C_A + k_I C_I + S_{reag}$$

Check out the Additional Resources at the end of the textbook for more information about linear regression with errors in both variables, curvilinear regression, and multivariate regression.

where k_I is the interferent's sensitivity and C_I is the interferent's concentration. Multivariate calibration curves can be prepared using standards that contain known amounts of both the analyte and the interferent, and modeled using multivariate regression.[11]

5E Blank Corrections

Thus far in our discussion of strategies for standardizing analytical methods, we have assumed the use of a suitable reagent blank to correct for signals arising from sources other than the analyte. We did not, however ask an important question—"What constitutes an appropriate reagent blank?" Surprisingly, the answer is not immediately obvious.

In one study, approximately 200 analytical chemists were asked to evaluate a data set consisting of a normal calibration curve, a separate analyte-free blank, and three samples of different size but drawn from the same source.[12] The first two columns in Table 5.3 shows a series of external standards and their corresponding signals. The normal calibration curve for the data is

$$S_{std} = 0.0750 \times W_{std} + 0.1250$$

where the y-intercept of 0.1250 is the calibration blank. A separate reagent blank gives the signal for an analyte-free sample.

In working up this data, the analytical chemists used at least four different approaches for correcting signals: (a) ignoring both the calibration blank, CB, and the reagent blank, RB, which clearly is incorrect; (b) using the calibration blank only; (c) using the reagent blank only; and (d) using both the calibration blank and the reagent blank. Table 5.4 shows the equa-

11 Beebe, K. R.; Kowalski, B. R. *Anal. Chem.* **1987**, *59*, 1007A–1017A.
12 Cardone, M. J. *Anal. Chem.* **1986**, *58*, 433–438.

Table 5.3 Data Used to Study the Blank in an Analytical Method

W_{std}	S_{std}	Sample Number	W_{samp}	S_{samp}
1.6667	0.2500	1	62.4746	0.8000
5.0000	0.5000	2	82.7915	1.0000
8.3333	0.7500	3	103.1085	1.2000
11.6667	0.8413			
18.1600	1.4870		reagent blank	0.1000
19.9333	1.6200			

Calibration equation: $S_{std} = 0.0750 \times W_{std} + 0.1250$

W_{std}: weight of analyte used to prepare the external standard; diluted to volume, V.
W_{samp}: weight of sample used to prepare sample; diluted to volume, V.

Table 5.4 Equations and Resulting Concentrations of Analyte for Different Approaches to Correcting for the Blank

Approach for Correcting Signal	Equation	Concentration of Analyte in... Sample 1	Sample 2	Sample 3
ignore calibration and reagent blank	$C_A = \frac{W_A}{W_{samp}} = \frac{S_{samp}}{k_A W_{samp}}$	0.1707	0.1610	0.1552
use calibration blank only	$C_A = \frac{W_A}{W_{samp}} = \frac{S_{samp} - CB}{k_A W_{samp}}$	0.1441	0.1409	0.1390
use reagent blank only	$C_A = \frac{W_A}{W_{samp}} = \frac{S_{samp} - RB}{k_A W_{samp}}$	0.1494	0.1449	0.1422
use both calibration and reagent blank	$C_A = \frac{W_A}{W_{samp}} = \frac{S_{samp} - CB - RB}{k_A W_{samp}}$	0.1227	0.1248	0.1261
use total Youden blank	$C_A = \frac{W_A}{W_{samp}} = \frac{S_{samp} - TYB}{k_A W_{samp}}$	0.1313	0.1313	0.1313

C_A = concentration of analyte; W_A = weight of analyte; W_{samp} = weight of sample; k_A = slope of calibration curve (0.075–see Table 5.3); CB = calibration blank (0.125–see Table 5.3); RB = reagent blank (0.100–see Table 5.3); TYB = total Youden blank (0.185–see text)

tions for calculating the analyte's concentration using each approach, along with the resulting concentration for the analyte in each sample.

That all four methods give a different result for the analyte's concentration underscores the importance of choosing a proper blank, but does not tell us which blank is correct. Because all four methods fail to predict the same concentration of analyte for each sample, none of these blank corrections properly accounts for an underlying constant source of determinate error.

To correct for a constant method error, a blank must account for signals from any reagents and solvents used in the analysis, as well as any bias resulting from interactions between the analyte and the sample's matrix. Both the calibration blank and the reagent blank compensate for signals from reagents and solvents. Any difference in their values is due to indeterminate errors in preparing and analyzing the standards.

Unfortunately, neither a calibration blank nor a reagent blank can correct for a bias resulting from an interaction between the analyte and the sample's matrix. To be effective, the blank must include both the sample's matrix and the analyte and, consequently, must be determined using the sample itself. One approach is to measure the signal for samples of different size, and to determine the regression line for a plot of S_{samp} versus the

Because we are considering a matrix effect of sorts, you might think that the method of standard additions is one way to overcome this problem. Although the method of standard additions can compensate for proportional determinate errors, it cannot correct for a constant determinate error; see Ellison, S. L. R.; Thompson, M. T. "Standard additions: myth and reality," *Analyst*, **2008**, *133*, 992–997.

amount of sample. The resulting y-intercept gives the signal in the absence of sample, and is known as the TOTAL YOUDEN BLANK.[13] This is the true blank correction. The regression line for the three samples in Table 5.3 is

$$S_{samp} = 0.009844 \times W_{samp} + 0.185$$

giving a true blank correction of 0.185. As shown by the last row of Table 5.4, using this value to correct S_{samp} gives identical values for the concentration of analyte in all three samples.

The use of the total Youden blank is not common in analytical work, with most chemists relying on a calibration blank when using a calibration curve, and a reagent blank when using a single-point standardization. As long we can ignore any constant bias due to interactions between the analyte and the sample's matrix, which is often the case, the accuracy of an analytical method will not suffer. It is a good idea, however, to check for constant sources of error before relying on either a calibration blank or a reagent blank.

5F Using Excel and R for a Regression Analysis

Although the calculations in this chapter are relatively straightforward—consisting, as they do, mostly of summations—it can be quite tedious to work through problems using nothing more than a calculator. Both Excel and R include functions for completing a linear regression analysis and for visually evaluating the resulting model.

5F.1 Excel

Let's use Excel to fit the following straight-line model to the data in Example 5.9.

$$y = \beta_0 + \beta_1 x$$

	A	B
1	Cstd	Sstd
2	0.000	0.00
3	0.100	12.36
4	0.200	24.83
5	0.300	35.91
6	0.400	48.79
7	0.500	60.42

Figure 5.15 Portion of a spreadsheet containing data from Example 5.9 (Cstd = C_{std}; Sstd = S_{std}).

Enter the data into a spreadsheet, as shown in Figure 5.15. Depending upon your needs, there are many ways that you can use Excel to complete a linear regression analysis. We will consider three approaches here.

Use Excel's Built-In Functions

If all you need are values for the slope, β_1, and the y-intercept, β_0, you can use the following functions:

=intercept(*known_y's*, *known_x's*)

=slope(*known_y's*, *known_x's*)

where *known_y's* is the range of cells containing the signals (y), and *known_x's* is the range of cells containing the concentrations (x). For example, clicking on an empty cell and entering

13 Cardone, M. J. *Anal. Chem.* **1986**, *58*, 438–445.

=slope(B2:B7, A2:A7)

returns Excel's exact calculation for the slope (120.705 714 3).

Use Excel's Data Analysis Tools

To obtain the slope and the y-intercept, along with additional statistical details, you can use the data analysis tools in the Analysis ToolPak. The ToolPak is not a standard part of Excel's instillation. To see if you have access to the Analysis ToolPak on your computer, select **Tools** from the menu bar and look for the **Data Analysis...** option. If you do not see **Data Analysis...**, select **Add-ins...** from the **Tools** menu. Check the box for the **Analysis ToolPak** and click on **OK** to install them.

Once you install the Analysis ToolPak, it will continue to load each time you launch Excel.

Select **Data Analysis...** from the **Tools** menu, which opens the *Data Analysis* window. Scroll through the window, select **Regression** from the available options, and press **OK**. Place the cursor in the box for *Input Y range* and then click and drag over cells B1:B7. Place the cursor in the box for *Input X range* and click and drag over cells A1:A7. Because cells A1 and B1 contain labels, check the box for *Labels*. Select the radio button for *Output range* and click on any empty cell; this is where Excel will place the results. Clicking **OK** generates the information shown in Figure 5.16.

Including labels is a good idea. Excel's summary output uses the x-axis label to identify the slope.

There are three parts to Excel's summary of a regression analysis. At the top of Figure 5.16 is a table of *Regression Statistics*. The *standard error* is the standard deviation about the regression, s_r. Also of interest is the value for *Multiple R*, which is the model's correlation coefficient, r, a term with which you may already by familiar. The correlation coefficient is a measure of the extent to which the regression model explains the variation in y. Values of r range from –1 to +1. The closer the correlation coefficient is to ±1, the better the model is at explaining the data. A correlation coefficient of 0 means that there is no relationship between x and y. In developing the calculations for linear regression, we did not consider the correlation coefficient. There

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.99987244
R Square	0.9997449
Adjusted R Square	0.99968113
Standard Error	0.40329713
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2549.727156	2549.72716	15676.296	2.4405E-08
Residual	4	0.650594286	0.16264857		
Total	5	2550.37775			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.20857143	0.29188503	0.71456706	0.51436267	-0.60183133	1.01897419	-0.60183133	1.01897419
Cstd	120.705714	0.964064525	125.205016	2.4405E-08	118.029042	123.382387	118.029042	123.382387

Figure 5.16 Output from Excel's Regression command in the Analysis ToolPak. See the text for a discussion of how to interpret the information in these tables.

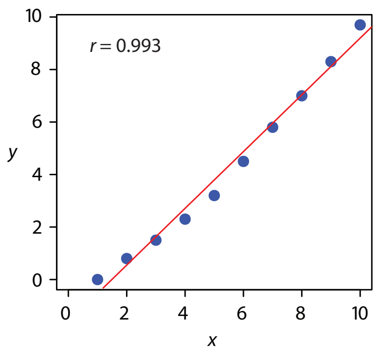


Figure 5.17 Example of fitting a straight-line to curvilinear data.

is a reason for this. For most straight-line calibration curves the correlation coefficient will be very close to +1, typically 0.99 or better. There is a tendency, however, to put too much faith in the correlation coefficient's significance, and to assume that an *r* greater than 0.99 means the linear regression model is appropriate. Figure 5.17 provides a counterexample. Although the regression line has a correlation coefficient of 0.993, the data clearly shows evidence of being curvilinear. The take-home lesson here is: don't fall in love with the correlation coefficient!

The second table in Figure 5.16 is entitled *ANOVA*, which stands for analysis of variance. We will take a closer look at ANOVA in Chapter 14. For now, it is sufficient to understand that this part of Excel's summary provides information on whether the linear regression model explains a significant portion of the variation in the values of y. The value for *F* is the result of an *F*-test of the following null and alternative hypotheses.

See Section 4F.2 and Section 4F.3 for a review of the F-test.

$$H_0\text{: regression model does not explain the variation in } y$$

$$H_A\text{: regression model does explain the variation in } y$$

The value in the column for *Significance F* is the probability for retaining the null hypothesis. In this example, the probability is $2.5\times10^{-6}\%$, suggesting that there is strong evidence for accepting the regression model. As is the case with the correlation coefficient, a small value for the probability is a likely outcome for any calibration curve, even when the model is inappropriate. The probability for retaining the null hypothesis for the data in Figure 5.17, for example, is $9.0\times10^{-7}\%$.

The third table in Figure 5.16 provides a summary of the model itself. The values for the model's coefficients—the slope, β_1, and the *y*-intercept, β_0—are identified as *intercept* and with your label for the *x*-axis data, which in this example is *Cstd*. The standard deviations for the coefficients, s_{b_0} and s_{b_1}, are in the column labeled *Standard error*. The column *t Stat* and the column *P-value* are for the following *t*-tests.

See Section 4F.1 for a review of the t-test.

$$\text{slope} \quad H_0\text{: } \beta_1 = 0,\ H_A\text{: } \beta_1 \neq 0$$

$$y\text{-intercept} \quad H_0\text{: } \beta_0 = 0,\ H_A\text{: } \beta_0 \neq 0$$

The results of these *t*-tests provide convincing evidence that the slope is not zero, but no evidence that the *y*-intercept significantly differs from zero. Also shown are the 95% confidence intervals for the slope and the *y*-intercept (*lower 95%* and *upper 95%*).

Program the Formulas Yourself

A third approach to completing a regression analysis is to program a spreadsheet using Excel's built-in formula for a summation

=sum(*first cell*:*last cell*)

and its ability to parse mathematical equations. The resulting spreadsheet is shown in Figure 5.18.

	A	B	C	D	E	F
1	x	y	xy	x^2	n =	6
2	0.000	0.00	=A2*B2	=A2^2	slope =	=(F1*C8 - A8*B8)/(F1*D8-A8^2)
3	0.100	12.36	=A3*B3	=A3^2	y-int =	=(B8-F2*A8)/F1
4	0.200	24.83	=A4*B4	=A4^2		
5	0.300	35.91	=A5*B5	=A5^2		
6	0.400	48.79	=A6*B6	=A6^2		
7	0.500	60.42	=A7*B7	=A7^2		
8						
9	=sum(A2:A7)	=sum(B2:B7)	=sum(C2:C7)	=sum(D2:D7)	<--sums	

Figure 5.18 Spreadsheet showing the formulas for calculating the slope and the y-intercept for the data in Example 5.9. The cells with the shading contain formulas that you must enter. Enter the formulas in cells C3 to C7, and cells D3 to D7. Next, enter the formulas for cells A9 to D9. Finally, enter the formulas in cells F2 and F3. When you enter a formula, Excel replaces it with the resulting calculation. The values in these cells should agree with the results in Example 5.9. You can simplify the entering of formulas by copying and pasting. For example, enter the formula in cell C2. Select **Edit: Copy**, click and drag your cursor over cells C3 to C7, and select **Edit: Paste**. Excel automatically updates the cell referencing.

Using Excel to Visualize the Regression Model

You can use Excel to examine your data and the regression line. Begin by plotting the data. Organize your data in two columns, placing the *x* values in the left-most column. Click and drag over the data and select **Insert: Chart...** from the main menu. This launches Excel's *Chart Wizard*. Select **xy-chart**, choosing the option without lines connecting the points. Click on **Next** and work your way through the screens, tailoring the plot to meet your needs. To add a regression line to the chart, click on the chart and select **Chart: Add Trendline...** from the main men. Pick the straight-line model and click **OK** to add the line to your chart. By default, Excel displays the regression line from your first point to your last point. Figure 5.19 shows the result for the data in Figure 5.15.

Excel's default options for *xy*-charts do not make for particularly attractive scientific figures. For example, Excel automatically adds grid lines parallel to the *x*-axis, which is a common practice in business charts. You can deselect them using the *Grid lines* tab in the *Chart Wizard*. Excel also defaults to a gray background. To remove this, just double-click on the chart's background and select *none* in the resulting pop-up window.

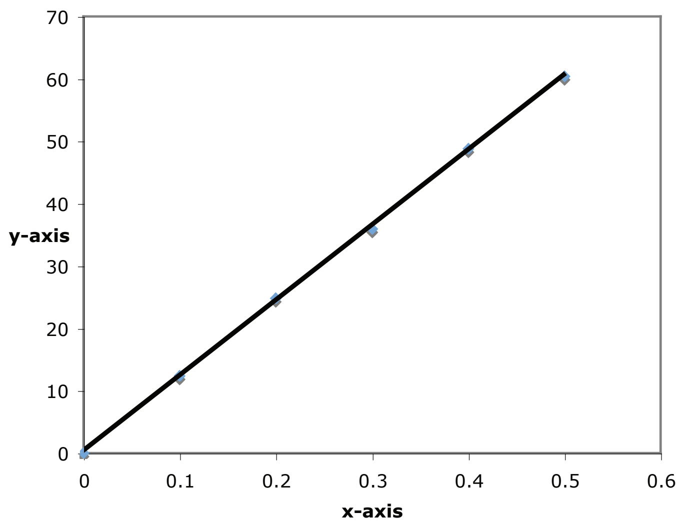


Figure 5.19 Example of an Excel scatterplot showing the data and a regression line.

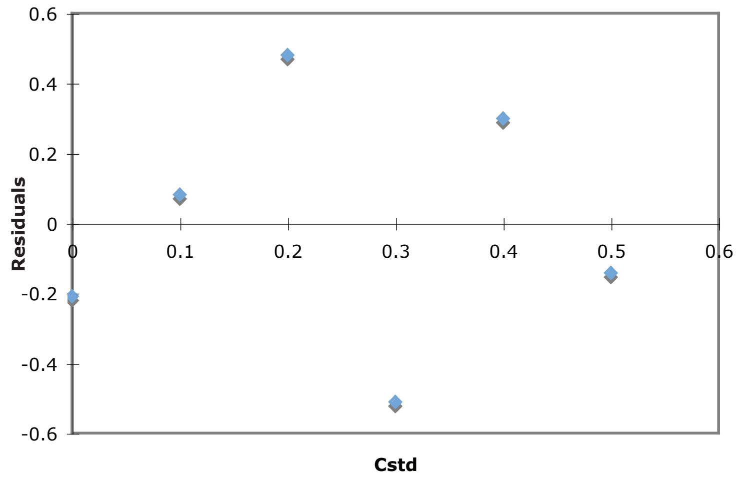


Figure 5.20 Example of Excel's plot of a regression model's residual errors.

Excel also will create a plot of the regression model's residual errors. To create the plot, build the regression model using the Analysis ToolPak, as described earlier. Clicking on the option for *Residual plots* creates the plot shown in Figure 5.20.

Limitations to Using Excel for a Regression Analysis

Excel's biggest limitation for a regression analysis is that it does not provide a function for calculating the uncertainty when predicting values of x. In terms of this chapter, Excel can not calculate the uncertainty for the analyte's concentration, C_A, given the signal for a sample, S_{samp}. Another limitation is that Excel does not include a built-in function for a weighted linear regression. You can, however, program a spreadsheet to handle these calculations.

> **Practice Exercise 5.6**
>
> Use Excel to complete the regression analysis in Practice Exercise 5.4.
>
> Click here to review your answer to this exercise.

5F.2 R

Let's use Excel to fit the following straight-line model to the data in Example 5.9.

$$y = \beta_0 + \beta_1 x$$

Entering Data and Creating the Regression Model

To begin, create objects containing the concentration of the standards and their corresponding signals.

```
> conc = c(0, 0.1, 0.2, 0.3, 0.4, 0.5)
> signal = c(0, 12.36, 24.83, 35.91, 48.79, 60.42)
```

The command for creating a straight-line linear regression model is

$$\text{lm}(y \sim x)$$

As you might guess, *lm* is short for linear model.

where *y* and *x* are the objects containing our data. To access the results of the regression analysis, we assign them to an object using the following command

> model = lm(signal ~ conc)

where *model* is the name we assign to the object.

You can choose any name for the object containing the results of the regression analysis.

Evaluating the Linear Regression Model

To evaluate the results of a linear regression we need to examine the data and the regression line, and to review a statistical summary of the model. To examine our data and the regression line, we use the **plot** command, which takes the following general form

plot(*x*, *y*, optional arguments to control style)

where *x* and *y* are objects containing our data, and the **abline** command

abline(*object*, optional arguments to control style)

where *object* is the object containing the results of the linear regression. Entering the commands

> plot(conc, signal, pch = 19, col = "blue", cex = 2)

> abline(model, col = "red")

creates the plot shown in Figure 5.21.

The name **abline** comes from the following common form for writing the equation of a straight-line.

$$y = a + bx$$

To review a statistical summary of the regression model, we use the **summary** command.

> summary(model)

The resulting output, shown in Figure 5.22, contains three sections.

The first section of R's summary of the regression model lists the residual errors. To examine a plot of the residual errors, use the command

> plot(model, which=1)

The reason for including the argument *which=1* is not immediately obvious. When you use R's *plot* command on an object created by the *lm* command, the default is to create four charts summarizing the model's suitability. The first of these charts is the residual plot; thus, *which=1* limits the output to this plot.

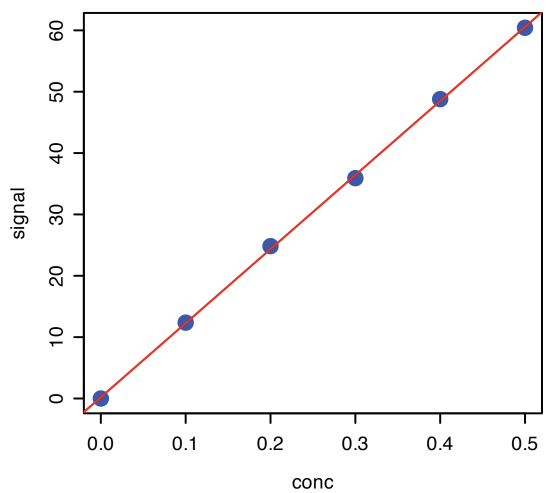


Figure 5.21 Example of a regression plot in R showing the data and the regression line. You can customize your plot by adjusting the plot command's optional arguments. The argument *pch* controls the symbol used for plotting points, the argument *col* allows you to select a color for the points or the line, and the argument *cex* sets the size for the points. You can use the command

help(plot)

to learn more about the options for plotting data in R.

```
> model=lm(signal~conc)
> summary(model)

Call:
lm(formula = signal ~ conc)

Residuals:
      1        2        3        4        5        6
-0.20857  0.08086  0.48029 -0.51029  0.29914 -0.14143

Coefficients:
            Estimate   Std. Error   t value   Pr(>|t|)
(Intercept)  0.2086     0.2919       0.715     0.514
conc       120.7057     0.9641     125.205     2.44e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4033 on 4 degrees of freedom
Multiple R-Squared: 0.9997,   Adjusted R-squared: 0.9997
F-statistic: 1.568e+04 on 1 and 4 DF,  p-value: 2.441e-08
```

Figure 5.22 The summary of R's regression analysis. See the text for a discussion of how to interpret the information in the output's three sections.

which produces the result shown in Figure 5.23. Note that R plots the residuals against the predicted (fitted) values of y instead of against the known values of x. The choice of how to plot the residuals is not critical, as you can see by comparing Figure 5.23 to Figure 5.20. The line in Figure 5.23 is a smoothed fit of the residuals.

The second section of Figure 5.22 provides the model's coefficients—the slope, β_1, and the y-intercept, β_0—along with their respective standard deviations (*Std. Error*). The column *t value* and the column *Pr(>|t|)* are for the following t-tests.

See Section 4F.1 for a review of the t-test.

$$\text{slope} \quad H_0\text{: } \beta_1 = 0,\ H_A\text{: } \beta_1 \neq 0$$

$$y\text{-intercept} \quad H_0\text{: } \beta_0 = 0,\ H_A\text{: } \beta_0 \neq 0$$

The results of these t-tests provide convincing evidence that the slope is not zero, but no evidence that the y-intercept significantly differs from zero.

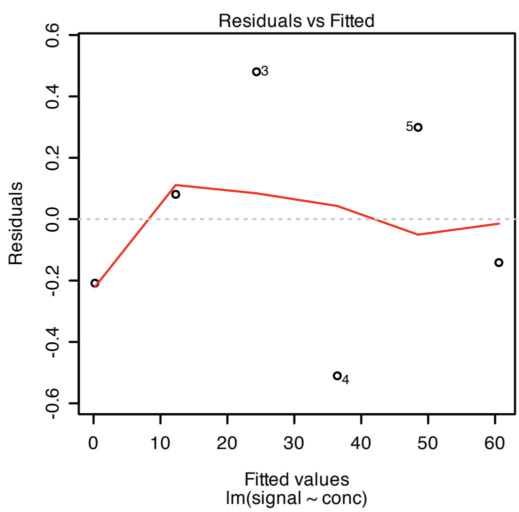


Figure 5.23 Example showing R's plot of a regression model's residual error.

The last section of the regression summary provides the standard deviation about the regression (*residual standard error*), the square of the correlation coefficient (*multiple R-squared*), and the result of an *F*-test on the model's ability to explain the variation in the y values. For a discussion of the correlation coefficient and the F-test of a regression model, as well as their limitations, refer to the section on using Excel's data analysis tools.

See Section 4F.2 and Section 4F.3 for a review of the *F*-test.

Predicting the Uncertainty in C_A Given S_{samp}

Unlike Excel, R includes a command for predicting the uncertainty in an analyte's concentration, C_A, given the signal for a sample, S_{samp}. This command is not part of R's standard installation. To use the command you need to install the "chemCal" package by entering the following command (*note: you will need an internet connection to download the package*).

> install.packages("chemCal")

After installing the package, you will need to load the functions into R using the following command. (*note: you will need to do this step each time you begin a new R session as the package does not automatically load when you start R*).

You need to install a package once, but you need to load the package each time you plan to use it. There are ways to configure R so that it automatically loads certain packages; see *An Introduction to R* for more information (click here to view a PDF version of this document).

> library("chemCal")

The command for predicting the uncertainty in C_A is **inverse.predict**, which takes the following form for an unweighted linear regression

inverse.predict(*object*, *newdata*, alpha = *value*)

where *object* is the object containing the regression model's results, *newdata* is an object containing values for S_{samp}, and *value* is the numerical value for the significance level. Let's use this command to complete Example 5.11. First, we create an object containing the values of S_{samp}

> sample = c(29.32, 29.16, 29.51)

and then we complete the computation using the following command

> inverse.predict(model, sample, alpha = 0.05)

producing the result shown in Figure 5.24. The analyte's concentration, C_A, is given by the value *$Prediction*, and its standard deviation, s_{C_A}, is shown as *$`Standard Error`*. The value for *$Confidence* is the confidence interval, $\pm ts_{C_A}$, for the analyte's concentration, and *$`Confidence Limits`* provides the lower limit and upper limit for the confidence interval for C_A.

Using R for a Weighted Linear Regression

R's command for an unweighted linear regression also allows for a weighted linear regression by including an additional argument, *weights*, whose value is an object containing the weights.

lm(*y* ~ *x*, *weights* = *object*)

```
> inverse.predict(model, sample, alpha = 0.05)
$Prediction
[1] 0.2412597

$`Standard Error`
[1] 0.002363588

$Confidence
[1] 0.006562373

$`Confidence Limits`
[1] 0.2346974 0.2478221
```

Figure 5.24 Output from R's command for predicting the analyte's concentration, C_A, from the sample's signal, S_{samp}.

You may have noticed that this way of defining weights is different than that shown in equation 5.28. In deriving equations for a weighted linear regression, you can choose to normalize the sum of the weights to equal the number of points, or you can choose not to—the algorithm in R does not normalize the weights.

Let's use this command to complete Example 5.12. First, we need to create an object containing the weights, which in R are the reciprocals of the standard deviations in y, $(s_{y_i})^{-2}$. Using the data from Example 5.12, we enter

```
> syi=c(0.02, 0.02, 0.07, 0.13, 0.22, 0.33)
> w=1/syi^2
```

to create the object containing the weights. The commands

```
> modelw=lm(signal ~ conc, weights = w)
> summary(modelw)
```

generate the output shown in Figure 5.25. Any difference between the results shown here and the results shown in Example 5.12 are the result of round-off errors in our earlier calculations.

Practice Exercise 5.7

Use Excel to complete the regression analysis in Practice Exercise 5.4.

Click here to review your answer to this exercise.

```
> modelw=lm(signal~conc, weights = w)
> summary(modelw)

Call:
lm(formula = signal ~ conc, weights = w)

Residuals:
     1      2      3      4      5      6
-2.223  2.571  3.676 -7.129 -1.413 -2.864

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)   0.04446     0.08542     0.52      0.63
conc        122.64111     0.93590   131.04  2.03e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.639 on 4 degrees of freedom
Multiple R-Squared: 0.9998,   Adjusted R-squared: 0.9997
F-statistic: 1.717e+04 on 1 and 4 DF,  p-value: 2.034e-08
```

Figure 5.25 The summary of R's regression analysis for a weighted linear regression. The types of information shown here is identical to that for the unweighted linear regression in Figure 5.22.

5G Key Terms

external standard	internal standard	linear regression
matrix matching	method of standard additions	multiple-point standardization
normal calibration curve	primary standard	reagent grade
residual error	secondary standard	serial dilution
single-point standardization	standard deviation about the regression	total Youden blank
unweighted linear regression	weighted linear regression	

As you review this chapter, try to define a key term in your own words. Check your answer by clicking on the key term, which will take you to the page where it was first introduced. Clicking on the KEY TERM there, will bring you back to this page so that you can continue with another key term.

5H Chapter Summary

In a quantitative analysis we measure a signal, S_{total}, and calculate the amount of analyte, n_A or C_A, using one of the following equations.

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

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

To obtain an accurate result we must eliminate determinate errors affecting the signal, S_{total}, the method's sensitivity, k_A, and the signal due to the reagents, S_{reag}.

To ensure that we accurately measure S_{total}, we calibrate our equipment and instruments. To calibrate a balance, for example, we a standard weight of known mass. The manufacturer of an instrument usually suggests appropriate calibration standards and calibration methods.

To standardize an analytical method we determine its sensitivity. There are several standardization strategies, including external standards, the method of standard addition and internal standards. The most common strategy is a multiple-point external standardization, resulting in a normal calibration curve. We use the method of standard additions, in which known amounts of analyte are added to the sample, when the sample's matrix complicates the analysis. When it is difficult to reproducibly handle samples and standards, we may choose to add an internal standard.

Single-point standardizations are common, but are subject to greater uncertainty. Whenever possible, a multiple-point standardization is preferred, with results displayed as a calibration curve. A linear regression analysis can provide an equation for the standardization.

A reagent blank corrects for any contribution to the signal from the reagents used in the analysis. The most common reagent blank is one in which an analyte-free sample is taken through the analysis. When a simple reagent blank does not compensate for all constant sources of determinate error, other types of blanks, such as the total Youden blank, can be used.

5I Problems

1. Describe how you would use a serial dilution to prepare 100 mL each of a series of standards with concentrations of 1.00×10^{-5}, 1.00×10^{-4}, 1.00×10^{-3}, and 1.00×10^{-2} M from a 0.100 M stock solution. Calculate the uncertainty for each solution using a propagation of uncertainty, and compare to the uncertainty if you were to prepare each solution by a single dilution of the stock solution. You will find tolerances for different types of volumetric glassware and digital pipets in Table 4.2 and Table 4.3. Assume that the uncertainty in the stock solution's molarity is ± 0.002.

2. Three replicate determinations of S_{total} for a standard solution that is 10.0 ppm in analyte give values of 0.163, 0.157, and 0.161 (arbitrary units). The signal for the reagent blank is 0.002. Calculate the concentration of analyte in a sample with a signal of 0.118.

3. A 10.00-g sample containing an analyte is transferred to a 250-mL volumetric flask and diluted to volume. When a 10.00 mL aliquot of the resulting solution is diluted to 25.00 mL it gives signal of 0.235 (arbitrary units). A second 10.00-mL portion of the solution is spiked with 10.00 mL of a 1.00-ppm standard solution of the analyte and diluted to 25.00 mL. The signal for the spiked sample is 0.502. Calculate the weight percent of analyte in the original sample.

4. A 50.00 mL sample containing an analyte gives a signal of 11.5 (arbitrary units). A second 50 mL aliquot of the sample, which is spiked with 1.00 mL of a 10.0-ppm standard solution of the analyte, gives a signal of 23.1. What is the analyte's concentration in the original sample?

5. An appropriate standard additions calibration curve based on equation 5.10 places $S_{spike} \times (V_o + V_{std})$ on the y-axis and $C_{std} \times V_{std}$ on the x-axis. Clearly explain why you can not plot S_{spike} on the y-axis and $C_{std} \times [V_{std}/(V_o + V_{std})]$ on the x-axis. In addition, derive equations for the slope and y-intercept, and explain how you can determine the amount of analyte in a sample from the calibration curve.

6. A standard sample contains 10.0 mg/L of analyte and 15.0 mg/L of internal standard. Analysis of the sample gives signals for the analyte and internal standard of 0.155 and 0.233 (arbitrary units), respectively. Sufficient internal standard is added to a sample to make its concentration 15.0 mg/L Analysis of the sample yields signals for the analyte and internal standard of 0.274 and 0.198, respectively. Report the analyte's concentration in the sample.

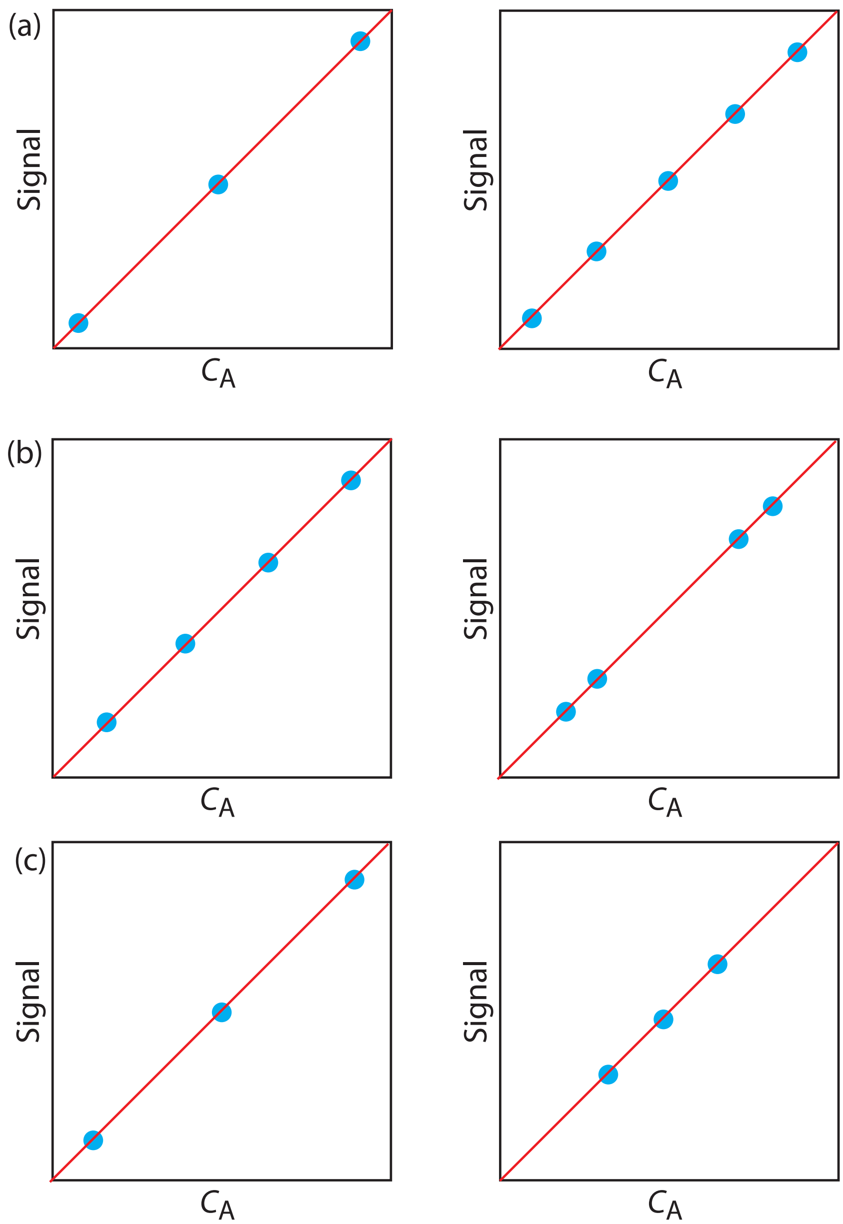


Figure 5.26 Calibration curves to accompany Problem 7.

7. For each of the pair of calibration curves shown in Figure 5.26, select the calibration curve using the more appropriate set of standards. Briefly explain the reasons for your selections. The scales for the *x*-axis and *y*-axis are the same for each pair.

8. The following data are for a series of external standards of Cd^{2+} buffered to a pH of 4.6.[14]

$[Cd^{2+}]$ (nM)	15.4	30.4	44.9	59.0	72.7	86.0
S_{total} (nA)	4.8	11.4	18.2	25.6	32.3	37.7

14 Wojciechowski, M.; Balcerzak, J. *Anal. Chim. Acta* **1991**, *249*, 433–445.

(a) Use a linear regression to determine the standardization relationship and report confidence intervals for the slope and the *y*-intercept.

(b) Construct a plot of the residuals and comment on their significance.

At a pH of 3.7 the following data were recorded for the same set of external standards.

$[Cd^{2+}]$ (nM)	15.4	30.4	44.9	59.0	72.7	86.0
S_{total} (nA)	15.0	42.7	58.5	77.0	101	118

(c) How much more or less sensitive is this method at the lower pH?

(d) A single sample is buffered to a pH of 3.7 and analyzed for cadmium, yielding a signal of 66.3. Report the concentration of Cd^{2+} in the sample and its 95% confidence interval.

9. To determine the concentration of analyte in a sample, a standard additions was performed. A 5.00-mL portion of sample was analyzed and then successive 0.10-mL spikes of a 600.0-mg/L standard of the analyte were added, analyzing after each spike. The following table shows the results of this analysis.

V_{spike} (mL)	0.00	0.10	0.20	0.30
S_{total} (arbitrary units)	0.119	0.231	0.339	0.442

Construct an appropriate standard additions calibration curve and use a linear regression analysis to determine the concentration of analyte in the original sample and its 95% confidence interval.

10. Troost and Olavsesn investigated the application of an internal standardization to the quantitative analysis of polynuclear aromatic hydrocarbons.[15] The following results were obtained for the analysis of phenanthrene using isotopically labeled phenanthrene as an internal standard. Each solution was analyzed twice.

C_A/C_{IS}	0.50	1.25	2.00	3.00	4.00
S_A/S_{IS}	0.514	0.993	1.486	2.044	2.342
	0.522	1.024	1.471	20.80	2.550

(a) Determine the standardization relationship using a linear regression, and report confidence intervals for the slope and the *y*-intercept. Average the replicate signals for each standard before completing the linear regression analysis.

(b) Based on your results explain why the authors concluded that the internal standardization was inappropriate.

15 Troost, J. R.; Olavesen, E. Y. *Anal. Chem.* **1996**, *68*, 708–711.

11. In Chapter 4 we used a paired t-test to compare two analytical methods used to independently analyze a series of samples of variable composition. An alternative approach is to plot the results for one method versus the results for the other method. If the two methods yield identical results, then the plot should have an expected slope, β_1, of 1.00 and an expected y-intercept, β_0, of 0.0. We can use a t-test to compare the slope and the y-intercept from a linear regression to the expected values. The appropriate test statistic for the y-intercept is found by rearranging equation 5.23.

$$t_{\text{exp}} = \frac{|\beta_0 - b_0|}{s_{b_0}} = \frac{|b_0|}{s_{b_0}}$$

Rearranging equation 5.22 gives the test statistic for the slope.

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

Reevaluate the data in problem 25 from Chapter 4 using the same significance level as in the original problem.

Although this is a common approach for comparing two analytical methods, it does violate one of the requirements for an unweighted linear regression—that indeterminate errors affect y only. Because indeterminate errors affect both analytical methods, the result of unweighted linear regression is biased. More specifically, the regression underestimates the slope, b_1, and overestimates the y-intercept, b_0. We can minimize the effect of this bias by placing the more precise analytical method on the x-axis, by using more samples to increase the degrees of freedom, and by using samples that uniformly cover the range of concentrations.

For more information, see Miller, J. C.; Miller, J. N. *Statistics for Analytical Chemistry*, 3rd ed. Ellis Horwood PTR Prentice-Hall: New York, 1993. Alternative approaches are found in Hartman, C.; Smeyers-Verbeke, J.; Penninckx, W.; Massart, D. L. *Anal. Chim. Acta* **1997**, *338*, 19–40, and Zwanziger, H. W.; Sârbu, C. *Anal. Chem.* **1998**, *70*, 1277–1280.

12. Consider the following three data sets, each containing value of y for the same values of x.

	Data Set 1	Data Set 2	Data Set 3
x	y_1	y_2	y_3
10.00	8.04	9.14	7.46
8.00	6.95	8.14	6.77
13.00	7.58	8.74	12.74
9.00	8.81	8.77	7.11
11.00	8.33	9.26	7.81
14.00	9.96	8.10	8.84
6.00	7.24	6.13	6.08
4.00	4.26	3.10	5.39
12.00	10.84	9.13	8.15
7.00	4.82	7.26	6.42
5.00	5.68	4.74	5.73

(a) An unweighted linear regression analysis for the three data sets gives nearly identical results. To three significant figures, each data set has a slope of 0.500 and a y-intercept of 3.00. The standard deviations in the slope and the y-intercept are 0.118 and 1.125 for each data set. All three standard deviations about the regression are 1.24, and all three data regression lines have a correlation coefficients of 0.816. Based on these results for a linear regression analysis, comment on the similarity of the data sets.

(b) Complete a linear regression analysis for each data set and verify that the results from part (a) are correct. Construct a residual plot for each data set. Do these plots change your conclusion from part (a)? Explain.

(c) Plot each data set along with the regression line and comment on your results.

(d) Data set 3 appears to contain an outlier. Remove this apparent outlier and reanalyze the data using a linear regression. Comment on your result.

(e) Briefly comment on the importance of visually examining your data.

13. Fanke and co-workers evaluated a standard additions method for a voltammetric determination of Tl.[16] A summary of their results is tabulated in the following table.

ppm Tl added	Instrument Response (μA)						
0.000	2.53	2.50	2.70	2.63	2.70	2.80	2.52
0.387	8.42	7.96	8.54	8.18	7.70	8.34	7.98
1.851	29.65	28.70	29.05	28.30	29.20	29.95	28.95
5.734	84.8	85.6	86.0	85.2	84.2	86.4	87.8

Use a weighted linear regression to determine the standardization relationship for this data.

5J Solutions to Practice Exercises

Practice Exercise 5.1

Substituting the sample's absorbance into the calibration equation and solving for C_A give

$$S_{samp} = 0.114 = 29.59 \text{ M}^{-1} \times C_A + 0.015$$
$$C_A = 3.35 \times 10^{-3} \text{ M}$$

For the one-point standardization, we first solve for k_A

16 Franke, J. P.; de Zeeuw, R. A.; Hakkert, R. *Anal. Chem.* **1978**, *50*, 1374–1380.

$$k_A = \frac{S_{std}}{C_{std}} = \frac{0.0931}{3.16 \times 10^{-3}\ \text{M}} = 29.46\ \text{M}^{-1}$$

and then use this value of k_A to solve for C_A.

$$C_A = \frac{S_{samp}}{k_A} = \frac{0.114}{29.46\ \text{M}^{-1}} = 3.87 \times 10^{-3}\ \text{M}$$

When using multiple standards, the indeterminate errors affecting the signal for one standard are partially compensated for by the indeterminate errors affecting the other standards. The standard selected for the one-point standardization has a signal that is smaller than that predicted by the regression equation, which underestimates k_A and overestimates C_A.

Click here to return to the chapter.

Practice Exercise 5.2

We begin with equation 5.8

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

rewriting it as

$$0 = \frac{k_A C_A V_o}{V_f} + k_A \times \left\{C_{std}\frac{V_{std}}{V_f}\right\}$$

which is in the form of the linear equation

$$Y = y\text{-intercept} + \text{slope} \times X$$

where Y is S_{spike} and X is $C_{std} \times V_{std}/V_f$. The slope of the line, therefore, is k_A, and the y-intercept is $k_A C_A V_o/V_f$. The x-intercept is the value of X when Y is zero, or

$$0 = \frac{k_A C_A V_o}{V_f} + k_A \times \{x\text{-intercept}\}$$

$$x\text{-intercept} = -\frac{k_A C_A V_o / V_f}{k_A} = -\frac{C_A V_o}{V_f}$$

Click here to return to the chapter.

Practice Exercise 5.3

Using the calibration equation from Figure 5.7a, we find that the x-intercept is

$$\text{x-intercept} = -\frac{0.1478}{0.0854 \text{ mL}^{-1}} = -1.731 \text{ mL}$$

Plugging this into the equation for the *x*-intercept and solving for C_A gives the concentration of Mn^{2+} as

$$\text{x-intercept} = -3.478 \text{ mL} = -\frac{C_A \times 25.00 \text{ mL}}{100.6 \text{ mg/L}} = 6.96 \text{ mg/L}$$

For Figure 7b, the *x*-intercept is

$$\text{x-intercept} = -\frac{0.1478}{0.0425 \text{ mL}^{-1}} = -3.478 \text{ mL}$$

and the concentration of Mn^{2+} is

$$\text{x-intercept} = -3.478 \text{ mL} = -\frac{C_A \times 25.00 \text{ mL}}{50.00 \text{ L}} = 6.96 \text{ mg/L}$$

Click here to return to the chapter.

Practice Exercise 5.4

We begin by setting up a table to help us organize the calculation.

x_i	y_i	x_iy_i	x_i^2
0.000	0.00	0.000	0.000
1.55×10^{-3}	0.050	7.750×10^{-5}	2.403×10^{-6}
3.16×10^{-3}	0.093	2.939×10^{-4}	9.986×10^{-6}
4.74×10^{-3}	0.143	6.778×10^{-4}	2.247×10^{-5}
6.34×10^{-3}	0.188	1.192×10^{-3}	4.020×10^{-5}
7.92×10^{-3}	0.236	1.869×10^{-3}	6.273×10^{-5}

Adding the values in each column gives

$$\sum_i x_i = 2.371 \times 10^{-2} \qquad \sum_i y_i = 0.710$$

$$\sum_i x_i y_i = 4.110 \times 10^{-3} \qquad \sum_i x_i^2 = 1.278 \times 10^{-4}$$

Substituting these values into equation 5.17 and equation 5.18, we find that the slope and the *y*-intercept are

$$b_1 = \frac{6 \times (4.110 \times 10^{-3}) - (2.371 \times 10^{-2}) \times (0.710)}{(6 \times 1.378 \times 10^{-4}) - (2.371 \times 10^{-2})^2} = 29.57$$

$$b_0 = \frac{0.710 - 29.57 \times (2.371 \times 10^{-2})}{6} = 0.0015$$

The regression equation is

$$S_{\text{std}} = 29.57 \times C_{\text{std}} + 0.0015$$

To calculate the 95% confidence intervals, we first need to determine the standard deviation about the regression. The following table will help us organize the calculation.

x_i	y_i	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$
0.000	0.00	0.0015	2.250×10^{-6}
1.55×10^{-3}	0.050	0.0473	7.110×10^{-6}
3.16×10^{-3}	0.093	0.0949	3.768×10^{-6}
4.74×10^{-3}	0.143	0.1417	1.791×10^{-6}
6.34×10^{-3}	0.188	0.1890	9.483×10^{-7}
7.92×10^{-3}	0.236	0.2357	9.339×10^{-8}

Adding together the data in the last column gives the numerator of equation 5.19 as 1.596×10^{-5}. The standard deviation about the regression, therefore, is

$$s_r = \sqrt{\frac{1.596 \times 10^{-6}}{6-2}} = 1.997 \times 10^{-3}$$

Next, we need to calculate the standard deviations for the slope and the y-intercept using equation 5.20 and equation 5.21.

$$s_{b_1} = \sqrt{\frac{6 \times (1.997 \times 10^{-3})^2}{6 \times (1.378 \times 10^{-4}) - (2.371 \times 10^{-2})^2}} = 0.3007$$

$$s_{b_0} = \sqrt{\frac{(1.997 \times 10^{-3})^2 \times (1.378 \times 10^{-4})}{6 \times (1.378 \times 10^{-4}) - (2.371 \times 10^{-2})^2}} = 1.441 \times 10^{-3}$$

The 95% confidence intervals are

$$\beta_1 = b_1 \pm ts_{b_1} = 29.57 \pm (2.78 \times 0.3007) = 29.57 \text{ M}^{-1} \pm 0.85 \text{ M}^{-1}$$

$$\beta_0 = b_0 \pm ts_{b_0} = 0.0015 \pm \{2.78 \times (1.441 \times 10^{-3}\} = 0.0015 \pm 0.0040$$

With an average S_{samp} of 0.114, the concentration of analyte, C_{A}, is

$$C_{\text{A}} = \frac{S_{\text{samp}} - b_0}{b_1} = \frac{0.114 - 0.0015}{29.57 \text{ M}^{-1}} = 3.80 \times 10^{-3} \text{ M}$$

The standard deviation in C_A is

$$s_{C_A} = \frac{1.997\times10^{-3}}{29.57}\sqrt{\frac{1}{3}+\frac{1}{6}+\frac{(0.114-0.1183)^2}{(29.57)^2\times(4.408\times10^{-5})}} = 4.778\times10^{-5}$$

and the 95% confidence interval is

$$\mu_{C_A} = C_A \pm ts_{C_A} = 3.80\times10^{-3} \pm \{2.78\times(4.778\times10^{-5})\}$$
$$= 3.80\times10^{-3}\text{ M} \pm 0.13\times10^{-3}\text{ M}$$

Click here to return to the chapter.

Practice Exercise 5.5

To create a residual plot, we need to calculate the residual error for each standard. The following table contains the relevant information.

x_i	y_i	$\hat{y}_i$	$y_i - \hat{y}_i$
0.000	0.00	0.0015	-0.0015
1.55×10^{-3}	0.050	0.0473	0.0027
3.16×10^{-3}	0.093	0.0949	-0.0019
4.74×10^{-3}	0.143	0.1417	0.0013
6.34×10^{-3}	0.188	0.1890	-0.0010
7.92×10^{-3}	0.236	0.2357	0.0003

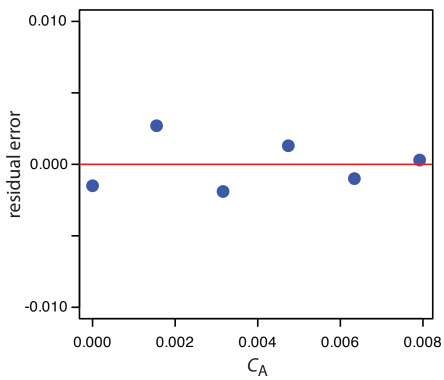


Figure 5.27 Plot of the residual errors for the data in Practice Exercise 5.5.

Figure 5.27 shows a plot of the resulting residual errors is shown here. The residual errors appear random and do not show any significant dependence on the analyte's concentration. Taken together, these observations suggest that our regression model is appropriate.

Click here to return to the chapter

Practice Exercise 5.6

Begin by entering the data into an Excel spreadsheet, following the format shown in Figure 5.15. Because Excel's Data Analysis tools provide most of the information we need, we will use it here. The resulting output, which is shown in Figure 5.28, contains the slope and the y-intercept, along with their respective 95% confidence intervals. Excel does not provide a function for calculating the uncertainty in the analyte's concentration, C_A, given the signal for a sample, S_{samp}. You must complete these calculations by hand. With an S_{samp}. of 0.114, C_A

$$C_A = \frac{S_{samp} - b_0}{b_1} = \frac{0.114 - 0.0014}{29.59\text{ M}^{-1}} = 3.80\times10^{-3}\text{ M}$$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.99979366
R Square	0.99958737
Adjusted R Sq	0.99948421
Standard Erroı	0.00199602
Observations	6

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.0386054	0.0386054	9689.9103	6.3858E-08
Residual	4	1.5936E-05	3.9841E-06		
Total	5	0.03862133			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.00139272	0.00144059	0.96677158	0.38840479	-0.00260699	0.00539242	-0.00260699	0.00539242
Cstd	29.5927329	0.30062507	98.437342	6.3858E-08	28.7580639	30.4274019	28.7580639	30.4274019

Figure 5.28 Excel's summary of the regression results for Practice Exercise 5.6.

The standard deviation in C_{A} is

$$s_{C_{\text{A}}} = \frac{1.996\times10^{-3}}{29.59}\sqrt{\frac{1}{3}+\frac{1}{6}+\frac{(0.114-0.1183)^2}{(29.59)^2\times(4.408\times10^{-5})}} = 4.772\times10^{-5}$$

and the 95% confidence interval is

$$\begin{aligned}\mu_{C_{\text{A}}} &= C_{\text{A}} \pm ts_{C_{\text{A}}} = 3.80\times10^{-3} \pm \{2.78\times(4.772\times10^{-5})\} \\ &= 3.80\times10^{-3}\text{ M} \pm 0.13\times10^{-3}\text{ M}\end{aligned}$$

Click here to return to the chapter

Practice Exercise 5.7

Figure 5.29 shows an R session for this problem, including loading the *chemCal* package, creating objects to hold the values for C_{std}, S_{std}, and *Ssamp*. Note that for *Ssamp*, we do not have the actual values for the three replicate measurements. In place of the actual measurements, we just enter the average signal three times. This is okay because the calculation depends on the average signal and the number of replicates, and not on the individual measurements.

Click here to return to the chapter

```
> library("chemCal")
> conc=c(0, 1.55e-3, 3.16e-3, 4.74e-3, 6.34e-3, 7.92e-3)
> signal=c(0, 0.050, 0.093, 0.143, 0.188, 0.236)
> model=lm(signal~conc)
> summary(model)

Call:
lm(formula = signal ~ conc)

Residuals:
      1          2          3          4          5          6
-0.0013927  0.0027385 -0.0019058  0.0013377 -0.0010106  0.0002328

Coefficients:
             Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)  0.001393   0.001441    0.967     0.388
conc        29.592733   0.300625   98.437     6.39e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001996 on 4 degrees of freedom
Multiple R-Squared: 0.9996,   Adjusted R-squared: 0.9995
F-statistic:  9690 on 1 and 4 DF,  p-value: 6.386e-08

> samp=c(0.114, 0.114, 0.114)
> inverse.predict(model,samp,alpha=0.05)
$Prediction
[1] 0.003805234

$`Standard Error`
[1] 4.771723e-05

$Confidence
[1] 0.0001324843

$`Confidence Limits`
[1] 0.003672750 0.003937719
```

Figure 5.29 R session for completing Practice Exercise 5.7.