between two signals, then the smallest detectable difference in the absolute amount or the relative amount of analyte is

$$\triangle n_A = \frac{\triangle S_A}{k_A} \quad \text{or} \quad \triangle C_A = \frac{\triangle S_A}{k_A}$$

Suppose, for example, that our analytical signal is a measurement of mass using a balance whose smallest detectable increment is ±0.0001 g. If our method's sensitivity is 0.200, then our method can conceivably detect a difference in mass of as little as

$$\triangle n_A = \frac{\pm 0.0001\,\text{g}}{0.200} = \pm 0.0005\,\text{g}$$

For two methods with the same $\triangle S_A$, the method with the greater sensitivity—that is, the method with the larger k_A—is better able to discriminate between smaller amounts of analyte.

3D.4 Specificity and Selectivity

An analytical method is specific if its signal depends only on the analyte.[4] Although SPECIFICITY is the ideal, few analytical methods are free from interferences. When an INTERFERENT contributes to the signal, we expand equation 3.1 and equation 3.2 to include its contribution to the sample's signal, S_{samp}

$$S_{samp} = S_A + S_I = k_A n_A + k_I n_I \qquad 3.3$$

$$S_{samp} = S_A + S_I = k_A C_A + k_I C_I \qquad 3.4$$

where S_I is the interferent's contribution to the signal, k_I is the interferent's sensitivity, and n_I and C_I are the moles (or grams) and the concentration of interferent in the sample, respectively.

SELECTIVITY is a measure of a method's freedom from interferences.[5] A method's selectivity for an interferent relative to the analyte is defined by a SELECTIVITY COEFFICIENT, $K_{A,I}$

$$K_{A,I} = \frac{k_I}{k_A} \qquad 3.5$$

which may be positive or negative depending on the sign of k_I and k_A. The selectivity coefficient is greater than $+1$ or less than -1 when the method is more selective for the interferent than for the analyte.

Determining the selectivity coefficient's value is easy if we already know the values for k_A and k_I. As shown by Example 3.1, we also can determine $K_{A,I}$ by measuring S_{samp} in the presence of and in the absence of the interferent.

Although k_A and k_I usually are positive, they can be negative. For example, some analytical methods work by measuring the concentration of a species that remains after is reacts with the analyte. As the analyte's concentration increases, the concentration of the species that produces the signal decreases, and the signal becomes smaller. If the signal in the absence of analyte is assigned a value of zero, then the subsequent signals are negative.

4   (a) Persson, B-A; Vessman, J. *Trends Anal. Chem.* **1998**, *17*, 117–119; (b) Persson, B-A; Vessman, J. *Trends Anal. Chem.* **2001**, *20*, 526–532.
5   Valcárcel, M.; Gomez-Hens, A.; Rubio, S. *Trends Anal. Chem.* **2001**, *20*, 386–393.

Example 3.1

A method for the analysis of Ca^{2+} in water suffers from an interference in the presence of Zn^{2+}. When the concentration of Ca^{2+} is 100 times greater than that of Zn^{2+}, an analysis for Ca^{2+} has a relative error of +0.5%. What is the selectivity coefficient for this method?

SOLUTION

Since only relative concentrations are reported, we can arbitrarily assign absolute concentrations. To make the calculations easy, we will let $C_{Ca} = 100$ (arbitrary units) and $C_{Zn} = 1$. A relative error of +0.5% means the signal in the presence of Zn^{2+} is 0.5% greater than the signal in the absence of Zn^{2+}. Again, we can assign values to make the calculation easier. If the signal for Cu^{2+} in the absence of Zn^{2+} is 100 (arbitrary units), then the signal in the presence of Zn^{2+} is 100.5.

The value of k_{Ca} is determined using equation 3.2

$$k_{Ca} = \frac{S_{Ca}}{C_{Ca}} = \frac{100}{100} = 1$$

In the presence of Zn^{2+} the signal is given by equation 3.4; thus

$$S_{samp} = 100.5 = k_{Ca}C_{Ca} + k_{Zn}C_{Zn} = (1 \times 100) + k_{Zn} \times 1$$

Solving for k_{Zn} gives its value as 0.5. The selectivity coefficient is

$$K_{Ca,Zn} = \frac{k_{Zn}}{k_{Ca}} = \frac{0.5}{1} = 0.5$$

If you are unsure why the signal in the presence of zinc is 100.5, note that the percentage relative error for this problem is given by

$$\frac{obtained\ result\ -\ 100}{100} \times$$

$$100 = +0.5\%$$

Solving gives an obtained result of 100.5.

Practice Exercise 3.1

Wang and colleagues describe a fluorescence method for the analysis of Ag^+ in water. When analyzing a solution that contains 1.0×10^{-9} M Ag^+ and 1.1×10^{-7} M Ni^{2+}, the fluorescence intensity (the signal) was +4.9% greater than that obtained for a sample of 1.0×10^{-9} M Ag^+. What is $K_{Ag,Ni}$ for this analytical method? The full citation for the data in this exercise is Wang, L.; Liang, A. N.; Chen, H.; Liu, Y.; Qian, B.; Fu, J. *Anal. Chim. Acta* **2008**, *616*, 170-176.

Click here to review your answer to this exercise.

A selectivity coefficient provides us with a useful way to evaluate an interferent's potential effect on an analysis. Solving equation 3.5 for k_I

$$k_I = K_{A,I} \times k_A \qquad 3.6$$

substituting in equation 3.3 and equation 3.4, and simplifying gives

$$S_{samp} = k_A\{n_A + K_{A,I} \times n_A\} \qquad 3.7$$

$$S_{samp} = k_A\{C_A + K_{A,I} \times C_I\} \qquad 3.8$$

An interferent will not pose a problem as long as the term $K_{A,I} \times n_I$ in equation 3.7 is significantly smaller than n_A, or if $K_{A,I} \times C_I$ in equation 3.8 is significantly smaller than C_A.

Example 3.2

Barnett and colleagues developed a method to determine the concentration of codeine in poppy plants.[6] As part of their study they evaluated the effect of several interferents. For example, the authors found that equimolar solutions of codeine and the interferent 6-methoxycodeine gave signals, respectively of 40 and 6 (arbitrary units).

(a) What is the selectivity coefficient for the interferent, 6-methoxycodeine, relative to that for the analyte, codeine.

(b) If we need to know the concentration of codeine with an accuracy of ±0.50%, what is the maximum relative concentration of 6-methoxycodeine that we can tolerate?

codeine

SOLUTION

(a) The signals due to the analyte, S_A, and the interferent, S_I, are

$$S_A = k_A C_A \qquad S_I = k_I C_I$$

Solving these equations for k_A and for k_I, and substituting into equation 3.6 gives

$$K_{A,I} = \frac{S_I/C_I}{S_A/C_A}$$

Because the concentrations of analyte and interferent are equimolar ($C_A = C_I$), the selectivity coefficient is

$$K_{A,I} = \frac{S_I}{S_A} = \frac{6}{40} = 0.15$$

(b) To achieve an accuracy of better than ±0.50% the term $K_{A,I} \times C_I$ in equation 3.8 must be less than 0.50% of C_A; thus

$$K_{A,I} \times C_I \leq 0.0050 \times C_A$$

Solving this inequality for the ratio C_I/C_A and substituting in the value for $K_{A,I}$ from part (a) gives

$$\frac{C_I}{C_A} \leq \frac{0.0050}{K_{A,I}} = \frac{0.0050}{0.15} = 0.033$$

Therefore, the concentration of 6-methoxycodeine must be less than 3.3% of codeine's concentration.
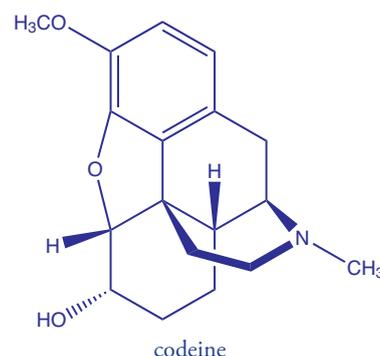
When a method's signal is the result of a chemical reaction—for example, when the signal is the mass of a precipitate—there is a good chance that the method is not very selective and that it is susceptible to an interference.

6   Barnett, N. W.; Bowser, T. A.; Geraldi, R. D.; Smith, B. *Anal. Chim. Acta* **1996**, *318*, 309–317.

> ## Practice Exercise 3.2
>
> Mercury (II) also is an interferent in the fluorescence method for Ag^+ developed by Wang and colleagues (see Practice Exercise 3.1 for the citation). The selectivity coefficient, $K_{Ag,Hg}$ has a value of -1.0×10^{-3}.
>
> (a) What is the significance of the selectivity coefficient's negative sign?
>
> (b) Suppose you plan to use this method to analyze solutions with concentrations of Ag^+ no smaller than 1.0 nM . What is the maximum concentration of Hg^{2+} you can tolerate if your percentage relative errors must be less than $\pm 1.0\%$?
>
> Click here to review your answers to this exercise.

Look back at Figure 1.1, which shows Fresenius' analytical method for the determination of nickel in ores. The reason there are so many steps in this procedure is that precipitation reactions generally are not very selective. The method in Figure 1.2 includes fewer steps because dimethylglyoxime is a more selective reagent. Even so, if an ore contains palladium, additional steps are needed to prevent the palladium from interfering.

Problems with selectivity also are more likely when the analyte is present at a very low concentration.[7]

3D.5 Robustness and Ruggedness

For a method to be useful it must provide reliable results. Unfortunately, methods are subject to a variety of chemical and physical interferences that contribute uncertainty to the analysis. If a method is relatively free from chemical interferences, we can use it to analyze an analyte in a wide variety of sample matrices. Such methods are considered ROBUST.

Random variations in experimental conditions introduces uncertainty. If a method's sensitivity, k, is too dependent on experimental conditions, such as temperature, acidity, or reaction time, then a slight change in any of these conditions may give a significantly different result. A RUGGED method is relatively insensitive to changes in experimental conditions.

3D.6 Scale of Operation

Another way to narrow the choice of methods is to consider three potential limitations: the amount of sample available for the analysis, the expected concentration of analyte in the samples, and the minimum amount of analyte that will produce a measurable signal. Collectively, these limitations define the analytical method's scale of operations.

We can display the scale of operations visually (Figure 3.5) by plotting the sample's size on the x-axis and the analyte's concentration on the y-axis.[8] For convenience, we divide samples into macro (>0.1 g), meso (10 mg–100 mg), micro (0.1 mg–10 mg), and ultramicro (<0.1 mg) sizes, and we divide analytes into major (>1% w/w), minor (0.01% w/w–1% w/w), trace (10^{-7}% w/w–0.01% w/w), and ultratrace (<10^{-7}% w/w) components. Together, the analyte's concentration and the sample's size provide a characteristic description for an analysis. For example, in a microtrace analysis the

7   Rodgers, L. B. *J. Chem. Educ.* **1986**, *63*, 3–6.

8   (a) Sandell, E. B.; Elving, P. J. in Kolthoff, I. M.; Elving, P. J., eds. *Treatise on Analytical Chemistry*, Interscience: New York, Part I, Vol. 1, Chapter 1, pp. 3–6; (b) Potts, L. W. *Quantitative Analysis–Theory and Practice*, Harper and Row: New York, 1987, pp. 12.

sample weighs between 0.1 mg and 10 mg and contains a concentration of analyte between 10^{-7}% w/w and 10^{-2}% w/w.

The diagonal lines connecting the axes show combinations of sample size and analyte concentration that contain the same absolute mass of analyte. As shown in Figure 3.5, for example, a 1-g sample that is 1% w/w analyte has the same amount of analyte (10 mg) as a 100-mg sample that is 10% w/w analyte, or a 10-mg sample that is 100% w/w analyte.

We can use Figure 3.5 to establish limits for analytical methods. If a method's minimum detectable signal is equivalent to 10 mg of analyte, then it is best suited to a major analyte in a macro or meso sample. Extending the method to an analyte with a concentration of 0.1% w/w requires a sample of 10 g, which rarely is practical due to the complications of carrying such a large amount of material through the analysis. On the other hand, a small sample that contains a trace amount of analyte places significant restrictions on an analysis. For example, a 1-mg sample that is 10^{-4}% w/w in analyte contains just 1 ng of analyte. If we isolate the analyte in 1 mL of solution, then we need an analytical method that reliably can detect it at a concentration of 1 ng/mL.

It should not surprise you to learn that a total analysis technique typically requires a macro or a meso sample that contains a major analyte. A concentration technique is particularly useful for a minor, trace, or ultratrace analyte in a macro, meso, or micro sample.
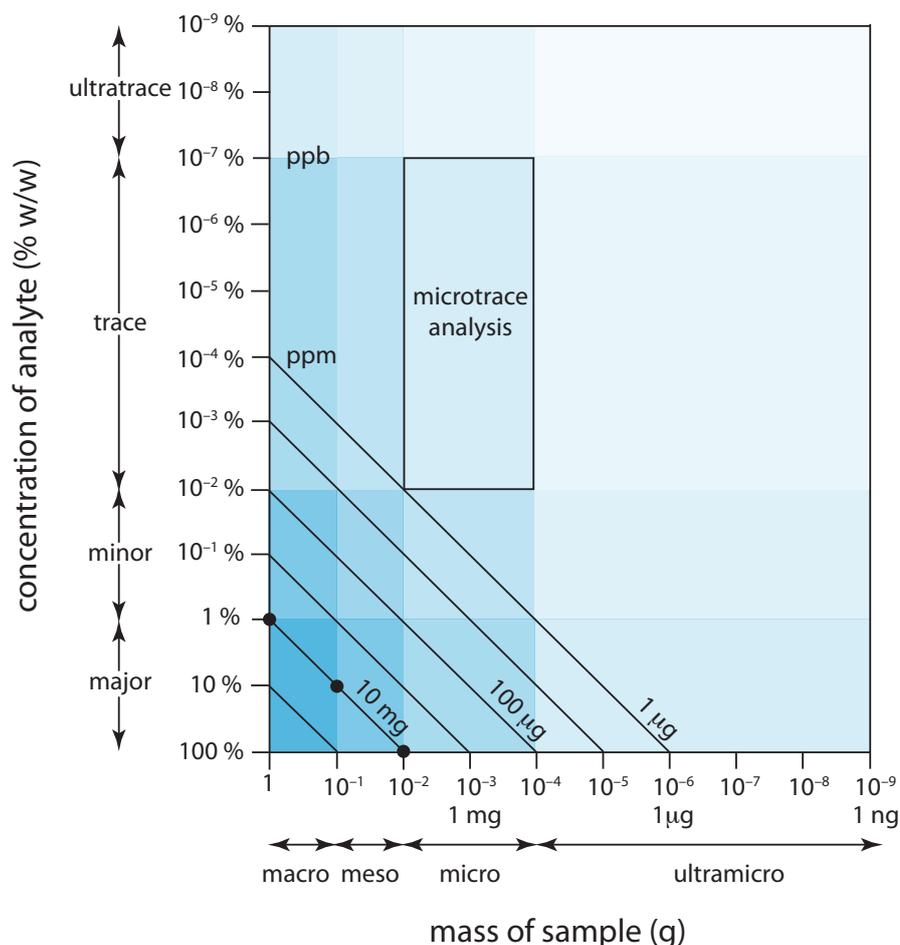

Figure 3.5 Scale of operations for analytical methods (adapted from references 8a and 8b).

The shaded areas define different types of analyses. The boxed area, for example, represents a microtrace analysis.

The diagonal lines show combinations of sample size and analyte concentration that contain the same mass of analyte. The three filled circles (●), for example, indicate analyses that use 10 mg of analyte.

3D.7 Equipment, Time, and Cost

Finally, we can compare analytical methods with respect to their equipment needs, the time needed to complete an analysis, and the cost per sample. Methods that rely on instrumentation are equipment-intensive and may require significant operator training. For example, the graphite furnace atomic absorption spectroscopic method for determining lead in water requires a significant capital investment in the instrument and an experienced operator to obtain reliable results. Other methods, such as titrimetry, require less expensive equipment and less training.

The time to complete an analysis for one sample often is fairly similar from method-to-method. This is somewhat misleading, however, because much of this time is spent preparing samples, preparing reagents, and gathering together equipment. Once the samples, reagents, and equipment are in place, the sampling rate may differ substantially. For example, it takes just a few minutes to analyze a single sample for lead using graphite furnace atomic absorption spectroscopy, but several hours to analyze the same sample using gravimetry. This is a significant factor in selecting a method for a laboratory that handles a high volume of samples.

The cost of an analysis depends on many factors, including the cost of equipment and reagents, the cost of hiring analysts, and the number of samples that can be processed per hour. In general, methods that rely on instruments cost more per sample then other methods.

3D.8 Making the Final Choice

Unfortunately, the design criteria discussed in this section are not mutually independent.[9] Working with smaller samples or improving selectivity often comes at the expense of precision. Minimizing cost and analysis time may decrease accuracy. Selecting a method requires carefully balancing the various design criteria. Usually, the most important design criterion is accuracy, and the best method is the one that gives the most accurate result. When the need for a result is urgent, as is often the case in clinical labs, analysis time may become the critical factor.

In some cases it is the sample's properties that determine the best method. A sample with a complex matrix, for example, may require a method with excellent selectivity to avoid interferences. Samples in which the analyte is present at a trace or ultratrace concentration usually require a concentration method. If the quantity of sample is limited, then the method must not require a large amount of sample.

Determining the concentration of lead in drinking water requires a method that can detect lead at the parts per billion concentration level. Selectivity is important because other metal ions are present at significantly higher concentrations. A method that uses graphite furnace atomic absorption spectroscopy is a common choice for determining lead in drinking

9   Valcárcel, M.; Ríos, A. *Anal. Chem.* **1993**, *65*, 781A–787A.

water because it meets these specifications. The same method is also useful for determining lead in blood where its ability to detect low concentrations of lead using a few microliters of sample is an important consideration.

3E   Developing the Procedure

After selecting a method, the next step is to develop a procedure that accomplish our goals for the analysis. In developing a procedure we give attention to compensating for interferences, to selecting and calibrating equipment, to acquiring a representative sample, and to validating the method.

3E.1   Compensating for Interferences

A method's accuracy depends on its selectivity for the analyte. Even the best method, however, may not be free from interferents that contribute to the measured signal. Potential interferents may be present in the sample itself or in the reagents used during the analysis.

When the sample is free of interferents, the total signal, S_{total}, is a sum of the signal due to the analyte, S_A, and the signal due to interferents in the reagents, S_{reag},

$$S_{total} = S_A + S_{reag} = k_A n_A + S_{reag} \qquad 3.9$$

$$S_{total} = S_A + S_{reag} = k_A C_A + S_{reag} \qquad 3.10$$

Without an independent determination of S_{reag} we cannot solve equation 3.9 or 3.10 for the moles or concentration of analyte.

To determine the contribution of S_{reag} in equations 3.9 and 3.10 we measure the signal for a METHOD BLANK, a solution that does not contain the sample. Consider, for example, a procedure in which we dissolve a 0.1-g sample in a portion of solvent, add several reagents, and dilute to 100 mL with additional solvent. To prepare the method blank we omit the sample and dilute the reagents to 100 mL using the solvent. Because the analyte is absent, S_{total} for the method blank is equal to S_{reag}. Knowing the value for S_{reag} makes it is easy to correct S_{total} for the reagent's contribution to the total signal; thus

$$(S_{total} - S_{reag}) = S_A = k_A n_A$$

$$(S_{total} - S_{reag}) = S_A = k_A C_A$$

By itself, a method blank cannot compensate for an interferent that is part of the sample's matrix. If we happen to know the interferent's identity and concentration, then we can be add it to the method blank; however, this is not a common circumstance and we must, instead, find a method for separating the analyte and interferent before continuing the analysis.

A method blank also is known as a reagent blank.

When the sample is a liquid, or is in solution, we use an equivalent volume of an inert solvent as a substitute for the sample.

3E.2 Calibration

A simple definition of a quantitative analytical method is that it is a mechanism for converting a measurement, the signal, into the amount of analyte in a sample. Assuming we can correct for interferents, a quantitative analysis is nothing more than solving equation 3.1 or equation 3.2 for n_A or for C_A.

To solve these equations we need the value of k_A. For a total analysis method usually we know the value of k_A because it is defined by the stoichiometry of the chemical reactions responsible for the signal. For a concentration method, however, the value of k_A usually is a complex function of experimental conditions. A CALIBRATION is the process of experimentally determining the value of k_A by measuring the signal for one or more standard samples, each of which contains a known concentration of analyte. With a single standard we can calculate the value of k_A using equation 3.1 or equation 3.2. When using several standards with different concentrations of analyte, the result is best viewed visually by plotting S_A versus the concentration of analyte in the standards. Such a plot is known as a CALIBRATION CURVE, an example of which is shown in Figure 3.6.

3E.3 Sampling

Selecting an appropriate method and executing it properly helps us ensure that our analysis is accurate. If we analyze the wrong sample, however, then the accuracy of our work is of little consequence.

A proper sampling strategy ensures that our samples are representative of the material from which they are taken. Biased or nonrepresentative sampling, and contaminating samples during or after their collection are two examples of sampling errors that can lead to a significant error in accuracy. It is important to realize that sampling errors are independent of errors in the analytical method. As a result, we cannot correct a sampling error in the laboratory by, for example, evaluating a reagent blank.

Figure 3.6 Example of a calibration curve. The filled circles (●) are the results for five standard samples, each with a different concentrations of analyte, and the line is the best fit to the data determined by a linear regression analysis. See Chapter 5 for a further discussion of calibration curves and an explanation of linear regression.
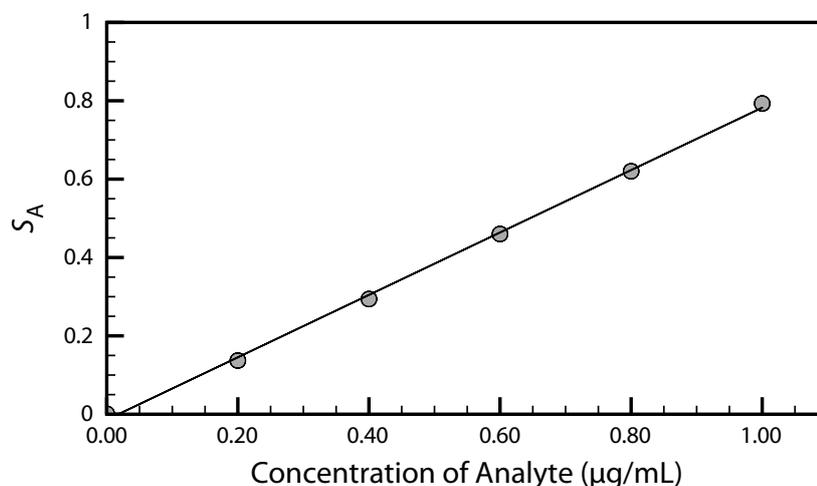
3E.4   Validation

If we are to have confidence in our procedure we must demonstrate that it can provide acceptable results, a process we call VALIDATION. Perhaps the most important part of validating a procedure is establishing that its precision and accuracy are appropriate for the problem we are trying to solve. We also ensure that the written procedure has sufficient detail so that different analysts or laboratories will obtain comparable results. Ideally, validation uses a standard sample whose composition closely matches the samples we will analyze. In the absence of appropriate standards, we can evaluate accuracy by comparing results to those obtained using a method of known accuracy.

You will find more details about validating analytical methods in Chapter 14.

3F   Protocols

Earlier we defined a protocol as a set of stringent written guidelines that specify an exact procedure that we must follow if an agency is to accept the results of our analysis. In addition to the considerations that went into the procedure's design, a protocol also contains explicit instructions regarding internal and external quality assurance and quality control (QA/QC) procedures.[10] The goal of internal QA/QC is to ensure that a laboratory's work is both accurate and precise. External QA/QC is a process in which an external agency certifies a laboratory.

As an example, let's outline a portion of the Environmental Protection Agency's protocol for determining trace metals in water by graphite furnace atomic absorption spectroscopy as part of its Contract Laboratory Program (CLP). The CLP protocol (see Figure 3.7) calls for an initial calibration using a method blank and three standards, one of which is at the detection limit. The resulting calibration curve is verified by analyzing initial calibration verification (ICV) and initial calibration blank (ICB) samples. The lab's result for the ICV sample must fall within ±10% of its expected concentration. If the result is outside this limit the analysis is stopped and the problem identified and corrected before continuing.

After a successful analysis of the ICV and ICB samples, the lab reverifies the calibration by analyzing a continuing calibration verification (CCV) sample and a continuing calibration blank (CCB). Results for the CCV also must be within ±10% of its expected concentration. Again, if the lab's result for the CCV is outside the established limits, the analysis is stopped, the problem identified and corrected, and the system recalibrated as described above. Additional CCV and the CCB samples are analyzed before the first sample and after the last sample, and between every set of ten samples. If the result for any CCV or CCB sample is unacceptable, the results for the last set of samples are discarded, the system is recalibrated, and the samples reanalyzed. By following this protocol, each result is bound by successful

10 (a) Amore, F. *Anal. Chem.* **1979**, *51*, 1105A–1110A; (b) Taylor, J. K. *Anal. Chem.* **1981**, *53*, 1588A–1593A.

Figure 3.7 Schematic diagram showing a portion of the EPA's protocol for determining trace metals in water using graphite furnace atomic absorption spectrometry.

The abbreviations are ICV: initial calibration verification; ICB: initial calibration blank; CCV: continuing calibration verification; CCB: continuing calibration blank.
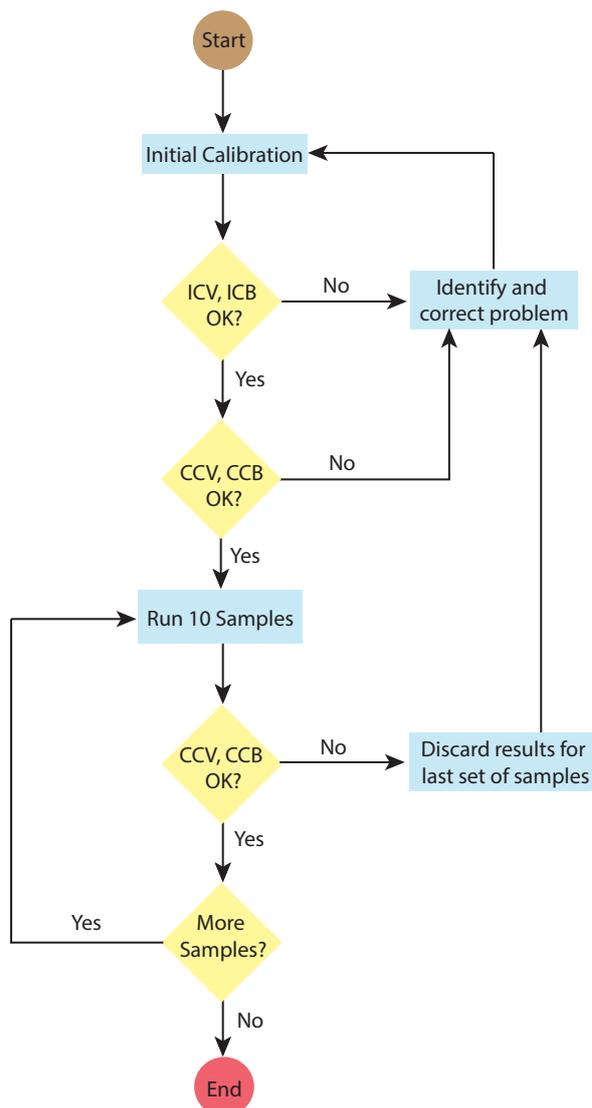
checks on the calibration. Although not shown in Figure 3.7, the protocol also contains instructions for analyzing duplicate or split samples, and for using spike tests to verify accuracy.

3G   The Importance of Analytical Methodology

The importance of the issues raised in this chapter is evident if we examine environmental monitoring programs. The purpose of a monitoring program is to determine the present status of an environmental system, and to assess long term trends in the system's health. These are broad and poorly defined goals. In many cases, an environmental monitoring program begins before the essential questions are known. This is not surprising since it is difficult to formulate questions in the absence of results. Without careful planning, however, a poor experimental design may result in data that has little value.

These concerns are illustrated by the Chesapeake Bay Monitoring Program. This research program, designed to study nutrients and toxic pollutants in the Chesapeake Bay, was initiated in 1984 as a cooperative venture between the federal government, the state governments of Maryland, Virginia, and Pennsylvania, and the District of Columbia. A 1989 review of the program highlights the problems common to many monitoring programs.[11]

At the beginning of the Chesapeake Bay monitoring program, little attention was given to selecting analytical methods, in large part because the eventual use of the data was not yet specified. The analytical methods initially chosen were standard methods already approved by the Environmental Protection Agency (EPA). In many cases these methods were not useful because they were designed to detect pollutants at their legally mandated maximum allowed concentrations. In unpolluted waters, however, the concentrations of these contaminants often are well below the detection limit of the EPA methods. For example, the detection limit for the EPA approved standard method for phosphate was 7.5 ppb. Since the actual phosphate concentrations in Chesapeake Bay were below the EPA method's detection limit, it provided no useful information. On the other hand, the detection limit for a non-approved variant of the EPA method, a method routinely used by chemical oceanographers, was 0.06 ppb, a more realistic detection limit for their samples. In other cases, such as the elemental analysis for particulate forms of carbon, nitrogen and phosphorous, EPA approved procedures provided poorer reproducibility than nonapproved methods.

3H   Key Terms

accuracy	analysis	analyte
calibration	calibration curve	concentration techniques
detection limit	determination	interferent
matrix	measurement	method
method blank	precision	procedure
protocol	QA/QC	robust
rugged	selectivity	selectivity coefficient
sensitivity	signal	specificity
technique	total analysis techniques	validation

3I   Chapter Summary

Every discipline has its own vocabulary and your success in studying analytical chemistry will improve if you master this vocabulary. Be sure you understand the difference between an analyte and its matrix, between a technique and a method, between a procedure and a protocol, and between a total analysis technique and a concentration technique.

---

11   D'Elia, C. F.; Sanders, J. G.; Capone, D. G. *Envrion. Sci. Technol.* **1989**, *23*, 768–774.

In selecting an analytical method we consider criteria such as accuracy, precision, sensitivity, selectivity, robustness, ruggedness, the amount of available sample, the amount of analyte in the sample, time, cost, and the availability of equipment. These criteria are not mutually independent, and often it is necessary to find an acceptable balance between them.

In developing a procedure or protocol, we give consideration to compensating for interferences, calibrating the method, obtaining an appropriate sample, and validating the analysis. Poorly designed procedures and protocols produce results that are insufficient to meet the needs of the analysis.

3J   Problems

1. When working with a solid sample, often it is necessary to bring the analyte into solution by digesting the sample with a suitable solvent. Any remaining solid impurities are removed by filtration before continuing with the analysis. In a typical total analysis method, the procedure might read

    > After digesting the sample in a beaker using approximately 25 mL of solvent, remove any solid impurities that remain by passing the solution the analyte through filter paper, collecting the filtrate in a clean Erlenmeyer flask. Rinse the beaker with several small portions of solvent, passing these rinsings through the filter paper and collecting them in the same Erlenmeyer flask. Finally, rinse the filter paper with several portions of solvent, collecting the rinsings in the same Erlenmeyer flask.

    For a typical concentration method, however, the procedure might state

    > After digesting the sample in a beaker using 25.00 mL of solvent, remove any solid impurities by filtering a portion of the solution containing the analyte. Collect and discard the first several mL of filtrate before collecting a sample of 5.00 mL for further analysis.

    Explain why these two procedures are different.

2. A certain concentration method works best when the analyte's concentration is approximately 10 ppb.

    (a) If the method requires a sample of 0.5 mL, about what mass of analyte is being measured?

    (b) If the analyte is present at 10% w/v, how would you prepare the sample for analysis?

    (c) Repeat for the case where the analyte is present at 10% w/w.

(d) Based on your answers to parts (a)–(c), comment on the method's suitability for the determination of a major analyte.

3. An analyst needs to evaluate the potential effect of an interferent, I, on the quantitative analysis for an analyte, A. She begins by measuring the signal for a sample in which the interferent is absent and the analyte is present with a concentration of 15 ppm, obtaining an average signal of 23.3 (arbitrary units). When she analyzes a sample in which the analyte is absent and the interferent is present with a concentration of 25 ppm, she obtains an average signal of 13.7.

(a) What is the sensitivity for the analyte?

(b) What is the sensitivity for the interferent?

(c) What is the value of the selectivity coefficient?

(d) Is the method more selective for the analyte or the interferent?

(e) What is the maximum concentration of interferent relative to that of the analyte if the error in the analysis is to be less than 1%?

4. A sample is analyzed to determine the concentration of an analyte. Under the conditions of the analysis the sensitivity is 17.2 ppm^{-1}. What is the analyte's concentration if S_{total} is 35.2 and S_{reag} is 0.6?

5. A method for the analysis of Ca^{2+} in water suffers from an interference in the presence of Zn^{2+}. When the concentration of Ca^{2+} is 50 times greater than that of Zn^{2+}, an analysis for Ca^{2+} gives a relative error of –2.0%. What is the value of the selectivity coefficient for this method?

6. The quantitative analysis for reduced glutathione in blood is complicated by many potential interferents. In one study, when analyzing a solution of 10.0 ppb glutathione and 1.5 ppb ascorbic acid, the signal was 5.43 times greater than that obtained for the analysis of 10.0 ppb glutathione.[12] What is the selectivity coefficient for this analysis? The same study found that analyzing a solution of 3.5×10^2 ppb methionine and 10.0 ppb glutathione gives a signal that is 0.906 times less than that obtained for the analysis of 10.0 ppb glutathione. What is the selectivity coefficient for this analysis? In what ways do these interferents behave differently?

7. Oungpipat and Alexander described a method for determining the concentration of glycolic acid (GA) in a variety of samples, including physiological fluids such as urine.[13] In the presence of only GA, the signal is

12 Jiménez-Prieto, R.; Velasco, A.; Silva, M; Pérez-Bendito, D. *Anal. Chem. Acta* **1992**, *269*, 273–279.
13 Oungpipat, W.; Alexander, P. W. *Anal. Chim. Acta* **1994**, *295*, 36–46.

$$S_{samp,1} = k_{GA}C_{GA}$$

and in the presence of both glycolic acid and ascorbic acid (AA), the signal is

$$S_{samp,2} = k_{GA}C_{GA} + k_{AA}C_{AA}$$

When the concentration of glycolic acid is 1.0×10^{-4} M and the concentration of ascorbic acid is 1.0×10^{-5} M, the ratio of their signals is

$$\frac{S_{samp,1}}{S_{samp,2}} = 1.44$$

(a) Using the ratio of the two signals, determine the value of the selectivity ratio $K_{GA,AA}$.

(b) Is the method more selective toward glycolic acid or ascorbic acid?

(c) If the concentration of ascorbic acid is 1.0×10^{-5} M, what is the smallest concentration of glycolic acid that can be determined such that the error introduced by failing to account for the signal from ascorbic acid is less than 1%?

8. Ibrahim and co-workers developed a new method for the quantitative analysis of hypoxanthine, a natural compound of some nucleic acids.[14] As part of their study they evaluated the method's selectivity for hypoxanthine in the presence of several possible interferents, including ascorbic acid.

(a) When analyzing a solution of 1.12×10^{-6} M hypoxanthine the authors obtained a signal of 7.45×10^{-5} amps. What is the sensitivity for hypoxanthine? You may assume the signal has been corrected for the method blank.

(b) When a solution containing 1.12×10^{-6} M hypoxanthine and 6.5×10^{-5} M ascorbic acid is analyzed a signal of 4.04×10^{-5} amps is obtained. What is the selectivity coefficient for this method?

(c) Is the method more selective for hypoxanthine or for ascorbic acid?

(d) What is the largest concentration of ascorbic acid that may be present if a concentration of 1.12×10^{-6} M hypoxanthine is to be determined within 1.0%?

9. Examine a procedure from *Standard Methods for the Analysis of Waters and Wastewaters* (or another manual of standard analytical methods) and identify the steps taken to compensate for interferences, to cali-

14 Ibrahim, M. S.; Ahmad, M. E.; Temerk, Y. M.; Kaucke, A. M. *Anal. Chim. Acta* **1996**, *328*, 47–52.

brate equipment and instruments, to standardize the method, and to acquire a representative sample.

3K   Solutions to Practice Exercises

Practice Exercise 3.1

Because the signal for Ag^+ in the presence of Ni^{2+} is reported as a relative error, we will assign a value of 100 as the signal for 1×10^{-9} M Ag^+. With a relative error of $+4.9\%$, the signal for the solution of 1×10^{-9} M Ag^+ and 1.1×10^{-7} M Ni^{2+} is 104.9. The sensitivity for Ag^+ is determined using the solution that does not contain Ni^{2+}; thus

$$k_{Ag} = \frac{S_{Ag}}{C_{Ag}} = \frac{100}{1 \times 10^{-9} \text{ M}} = 1.0 \times 10^{11} \text{ M}^{-1}$$

Substituting into equation 3.4 values for k_{Ag}, S_{samp}, and the concentrations of Ag^+ and Ni^{2+}

$$104.9 = (1.0 \times 10^{11} \text{ M}^{-1}) \times (1.0 \times 10^{-9} \text{ M}) + k_{Ni} \times (1.1 \times 10^{-7} \text{ M})$$

and solving gives k_{Ni} as 4.5×10^7 M^{-1}. The selectivity coefficient is

$$K_{Ag,Ni} = \frac{k_{Ni}}{k_{Ag}} = \frac{4.5 \times 10^7 \text{ M}^{-1}}{1.0 \times 10^{11} \text{ M}^{-1}} = 4.5 \times 10^{-4}$$

Click here to return to the chapter.

Practice Exercise 3.2

(a) A negative value for $K_{Ag,Hg}$ means that the presence of Hg^{2+} decreases the signal from Ag^+.

(b) In this case we need to consider an error of -1%, since the effect of Hg^{2+} is to decrease the signal from Ag^+. To achieve this error, the term $K_{A,I} \times C_I$ in equation 3.8 must be less than -1% of C_A; thus

$$K_{Ag,Hg} \times C_{Hg} = -0.01 \times C_{Ag}$$

Substituting in known values for $K_{Ag,Hg}$ and C_{Ag}, we find that the maximum concentration of Hg^{2+} is 1.0×10^{-8} M.

Click here to return to the chapter.

Chapter 4

Evaluating Analytical Data

Chapter Overview

When we use an analytical method we make three separate evaluations of experimental error. First, before we begin the analysis we evaluate potential sources of errors to ensure they will not adversely effect our results. Second, during the analysis we monitor our measurements to ensure that errors remain acceptable. Finally, at the end of the analysis we evaluate the quality of the measurements and results, and compare them to our original design criteria. This chapter provides an introduction to sources of error, to evaluating errors in analytical measurements, and to the statistical analysis of data.

Figure 4.1 An uncirculated 2005 Lincoln head penny. The "D" below the date indicates that this penny was produced at the United States Mint at Denver, Colorado. Pennies produced at the Philadelphia Mint do not have a letter below the date. Source: United States Mint image (www.usmint.gov).

4A   Characterizing Measurements and Results

Let's begin by choosing a simple quantitative problem that requires a single measurement: What is the mass of a penny? You probably recognize that our statement of the problem is too broad. For example, are we interested in the mass of a United States penny or of a Canadian penny, or is the difference relevant? Because a penny's composition and size may differ from country to country, let's narrow our problem to pennies from the United States.

There are other concerns we might consider. For example, the United States Mint produces pennies at two locations (Figure 4.1). Because it seems unlikely that a penny's mass depends on where it is minted, we will ignore this concern. Another concern is whether the mass of a newly minted penny is different from the mass of a circulating penny. Because the answer this time is not obvious, let's further narrow our question and ask "What is the mass of a circulating United States Penny?"

A good way to begin our analysis is to gather some preliminary data. Table 4.1 shows masses for seven pennies collected from my change jar. In examining this data we see that our question does not have a simple answer. That is, we can not use the mass of a single penny to draw a specific conclusion about the mass of any other penny (although we might conclude that all pennies weigh at least 3 g). We can, however, characterize this data by reporting the spread of the individual measurements around a central value.

4A.1   Measures of Central Tendency

One way to characterize the data in Table 4.1 is to assume that the masses of individual pennies are scattered randomly around a central value that is the best estimate of a penny's expected, or "true" mass. There are two common ways to estimate central tendency: the mean and the median.

MEAN

The MEAN, $\overline{X}$, is the numerical average for a data set. We calculate the mean by dividing the sum of the individual values by the size of the data set

Table 4.1 Masses of Seven Circulating U. S. Pennies	
Penny	Mass (g)
1	3.080
2	3.094
3	3.107
4	3.056
5	3.112
6	3.174
7	3.198

$$\overline{X} = \frac{\sum\limits_{i=1}^{n} X_i}{n}$$

where X_i is the ith measurement, and n is the size of the data set.

Example 4.1

What is the mean for the data in Table 4.1?

SOLUTION

To calculate the mean we add together the results for all measurements

$3.080 + 3.094 + 3.107 + 3.056 + 3.112 + 3.174 + 3.198 = 21.821\,\text{g}$

and divide by the number of measurements

$$\overline{X} = \frac{21.821\,\text{g}}{7} = 3.117\,\text{g}$$

The mean is the most common estimate of central tendency. It is not a robust estimate, however, because a single extreme value—one much larger or much smaller than the remainder of the data— influences strongly the mean's value.[1] For example, if we accidently record the third penny's mass as 31.07 g instead of 3.107 g, the mean changes from 3.117 g to 7.112 g!

An estimate for a statistical parameter is robust if its value is not affected too much by an unusually large or an unusually small measurement.

MEDIAN

The MEDIAN, $\widetilde{X}$, is the middle value when we order our data from the smallest to the largest value. When the data has an odd number of values, the median is the middle value. For an even number of values, the median is the average of the $n/2$ and the $(n/2)+1$ values, where n is the size of the data set.

When $n = 5$, the median is the third value in the ordered data set; for $n = 6$, the median is the average of the third and fourth members of the ordered data set.

Example 4.2

What is the median for the data in Table 4.1?

SOLUTION

To determine the median we order the measurements from the smallest to the largest value

3.056   3.080   3.094   3.107   3.112   3.174   3.198

Because there are seven measurements, the median is the fourth value in the ordered data; thus, the median is 3.107 g.

As shown by Examples 4.1 and 4.2, the mean and the median provide similar estimates of central tendency when all measurements are comparable in magnitude. The median, however, is a more robust estimate of central tendency because it is less sensitive to measurements with extreme values.

1   Rousseeuw, P. J. *J. Chemom.* **1991**, *5*, 1–20.

For example, if we accidently record the third penny's mass as 31.07 g instead of 3.107 g, the median's value changes from 3.107 g to 3.112 g.

4A.2 Measures of Spread

If the mean or the median provides an estimate of a penny's expected mass, then the spread of individual measurements about the mean or median provides an estimate of the difference in mass among pennies or of the uncertainty in measuring mass with a balance. Although we often define the spread relative to a specific measure of central tendency, its magnitude is independent of the central value. Although shifting all measurements in the same direction by adding or subtracting a constant value changes the mean or median, it does not change the spread. There are three common measures of spread: the range, the standard deviation, and the variance.

RANGE

The RANGE, w, is the difference between a data set's largest and smallest values.

$$w = X_{\text{largest}} - X_{\text{smallest}}$$

The range provides information about the total variability in the data set, but does not provide information about the distribution of individual values. The range for the data in Table 4.1 is

$$w = 3.198 \text{ g} - 3.056 \text{ g} = 0.142 \text{ g}$$

STANDARD DEVIATION

The STANDARD DEVIATION, s, describes the spread of individual values about their mean, and is given as

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}} \qquad 4.1$$

where X_i is one of n individual values in the data set, and $\overline{X}$ is the data set's mean value. Frequently, we report the relative standard deviation, s_r, instead of the absolute standard deviation.

$$s_r = \frac{s}{\overline{X}}$$

The percent relative standard deviation, $\%s_r$, is $s_r \times 100$.

<div style="border-left: 4px solid green; padding-left: 10px;">

Example 4.3

Report the standard deviation, the relative standard deviation, and the percent relative standard deviation for the data in Table 4.1?

SOLUTION

To calculate the standard deviation we first calculate the difference between each measurement and the data set's mean value (3.117), square the result-
</div>

As you might guess from this equation, the range is not a robust estimate of spread.

The relative standard deviation is important because it allows for a more meaningful comparison between data sets when the individual measurements differ significantly in magnitude. Consider again the data in Table 4.1. If we multiply each value by 10, the absolute standard deviation will increase by 10 as well; the relative standard deviation, however, is the same.

ing differences, and add them together to find the numerator of [equation 4.1](#).

$$(3.080 - 3.117)^2 = (-0.037)^2 = 0.001369$$
$$(3.094 - 3.117)^2 = (-0.023)^2 = 0.000529$$
$$(3.107 - 3.117)^2 = (-0.010)^2 = 0.000100$$
$$(3.056 - 3.117)^2 = (-0.061)^2 = 0.003721$$
$$(3.112 - 3.117)^2 = (-0.005)^2 = 0.000025$$
$$(3.174 - 3.117)^2 = (+0.057)^2 = 0.003249$$
$$(3.198 - 3.117)^2 = (+0.081)^2 = 0.006561$$
$$\overline{0.015554}$$

Next, we divide this sum of squares by $n-1$, where n is the number of measurements, and take the square root.

$$s = \sqrt{\frac{0.015554}{7-1}} = 0.051\,g$$

Finally, the relative standard deviation and percent relative standard deviation are

$$s_r = \frac{0.051\,g}{3.117\,g} = 0.016$$

$$\%s_r = (0.016) \times 100\% = 1.6\%$$

It is much easier to determine the standard deviation using a scientific calculator with built in statistical functions.

For obvious reasons, the numerator of [equation 4.1](#) is called a sum of squares.

Many scientific calculators include two keys for calculating the standard deviation. One key calculates the standard deviation for a data set of n samples drawn from a larger collection of possible samples, which corresponds to [equation 4.1](#). The other key calculates the standard deviation for all possible samples. The latter is known as the population's standard deviation, which we will cover later in this chapter. Your calculator's manual will help you determine the appropriate key for each.

Variance

Another common measure of spread is the VARIANCE, which is the square of the standard deviation. We usually report a data set's standard deviation, rather than its variance, because the mean value and the standard deviation share the same unit. As we will see shortly, the variance is a useful measure of spread because its values are additive.

Example 4.4

What is the variance for the data in [Table 4.1](#)?

SOLUTION

The variance is the square of the absolute standard deviation. Using the standard deviation from Example 4.3 gives the variance as

$$s^2 = (0.051)^2 = 0.0026$$

> ## Practice Exercise 4.1
>
> The following data were collected as part of a quality control study for the analysis of sodium in serum; results are concentrations of Na^+ in mmol/L.
>
> | 140 | 143 | 141 | 137 | 132 | 157 | 143 | 149 | 118 | 145 |
> |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
>
> Report the mean, the median, the range, the standard deviation, and the variance for this data. This data is a portion of a larger data set from Andrew, D. F.; Herzberg, A. M. *Data: A Collection of Problems for the Student and Research Worker*, Springer-Verlag:New York, 1985, pp. 151–155.
>
> Click <u>here</u> to review your answer to this exercise.

4B   Characterizing Experimental Errors

Characterizing a penny's mass using the data in <u>Table 4.1</u> suggests two questions. First, does our measure of central tendency agree with the penny's expected mass? Second, why is there so much variability in the individual results? The first of these questions addresses the accuracy of our measurements and the second addresses the precision of our measurements. In this section we consider the types of experimental errors that affect accuracy and precision.

4B.1   Errors That Affect Accuracy

The convention for representing a statistical parameter is to use a Roman letter for a value calculated from experimental data, and a Greek letter for its corresponding expected value. For example, <u>the</u> experimentally determined mean is $\overline{X}$, and its underlying expected value is μ. Likewise, the standard deviation by experiment is s, and the underlying expected value is σ.

Accuracy is how close a measure of central tendency is to its expected value, μ. We express accuracy either as an absolute error, e

$$e = \overline{X} - \mu \qquad 4.2$$

or as a percent relative error, $\%e_r$.

$$\%e_r = \frac{\overline{X} - \mu}{\mu} \times 100 \qquad 4.3$$

Although equation 4.2 and equation 4.3 use the mean as the measure of central tendency, we also can use the median.

We identify as determinate an error that affects the accuracy of an analysis. Each source of a DETERMINATE ERROR has a specific magnitude and sign. Some sources of determinate error are positive and others are negative, and some are larger in magnitude and others are smaller in magnitude. The cumulative effect of these determinate errors is a net positive or negative error in accuracy.

It is possible, although unlikely, that the positive and negative determinate errors will offset each other, producing a result with no net error in accuracy.

We assign determinate errors into four categories—sampling errors, method errors, measurement errors, and personal errors—each of which we consider in this section.

SAMPLING ERRORS

A determinate SAMPLING ERROR occurs when our sampling strategy does not provide a us with a representative sample. For example, if we monitor the environmental quality of a lake by sampling from a single site near a point source of pollution, such as an outlet for industrial effluent, then our results will be misleading. To determine the mass of a U. S. penny, our strategy for selecting pennies must ensure that we do not include pennies from other countries.

An awareness of potential sampling errors especially is important when we work with heterogeneous materials. Strategies for obtaining representative samples are covered in Chapter 5.

METHOD ERRORS

In any analysis the relationship between the signal, S_{total}, and the absolute amount of analyte, n_A, or the analyte's concentration, C_A, is

$$S_{total} = k_A n_A + S_{mb} \qquad 4.4$$

$$S_{total} = k_A C_A + S_{mb} \qquad 4.5$$

where k_A is the method's sensitivity for the analyte and S_{mb} is the signal from the method blank. A METHOD ERROR exists when our value for k_A or for S_{mb} is in error. For example, a method in which S_{total} is the mass of a precipitate assumes that k is defined by a pure precipitate of known stoichiometry. If this assumption is not true, then the resulting determination of n_A or C_A is inaccurate. We can minimize a determinate error in k_A by calibrating the method. A method error due to an interferent in the reagents is minimized by using a proper method blank.

MEASUREMENT ERRORS

The manufacturers of analytical instruments and equipment, such as glassware and balances, usually provide a statement of the item's maximum MEASUREMENT ERROR, or TOLERANCE. For example, a 10-mL volumetric pipet (Figure 4.2) has a tolerance of ±0.02 mL, which means the pipet delivers an actual volume within the range 9.98–10.02 mL at a temperature of 20 °C. Although we express this tolerance as a range, the error is determinate; that is, the pipet's expected volume, μ, is a fixed value within this stated range.

Volumetric glassware is categorized into classes based on its relative accuracy. Class A glassware is manufactured to comply with tolerances specified by an agency, such as the National Institute of Standards and Technology or the American Society for Testing and Materials. The tolerance level for Class A glassware is small enough that normally we can use it without calibration. The tolerance levels for Class B glassware usually are twice that for Class A glassware. Other types of volumetric glassware, such as beakers and graduated cylinders, are not used to measure volume accurately. Table 4.2 provides a summary of typical measurement errors for Class A volumetric glassware. Tolerances for digital pipets and for balances are provided in Table 4.3 and Table 4.4.

Figure 4.2 Close-up of a 10-mL volumetric pipet showing that it has a tolerance of ±0.02 mL at 20 °C.

We can minimize a determinate measurement error by calibrating our equipment. Balances are calibrated using a reference weight whose mass we can trace back to the SI standard kilogram. Volumetric glassware and digital pipets are calibrated by determining the mass of water delivered or contained and using the density of water to calculate the actual volume. It is never safe to assume that a calibration does not change during an analysis or over time. One study, for example, found that repeatedly exposing volumetric glassware to higher temperatures during machine washing and oven drying, led to small, but significant changes in the glassware's calibration.[2] Many instruments drift out of calibration over time and may require frequent recalibration during an analysis.

2   Castanheira, I.; Batista, E.; Valente, A.; Dias, G.; Mora, M.; Pinto, L.; Costa, H. S. *Food Control* **2006**, *17*, 719–726.

Table 4.2   Measurement Errors for Type A Volumetric Glassware[†]

Transfer Pipets		Volumetric Flasks		Burets	
Capacity (mL)	Tolerance (mL)	Capacity (mL)	Tolerance (mL)	Capacity (mL)	Tolerance (mL)
1	±0.006	5	±0.02	10	±0.02
2	±0.006	10	±0.02	25	±0.03
5	±0.01	25	±0.03	50	±0.05
10	±0.02	50	±0.05		
20	±0.03	100	±0.08		
25	±0.03	250	±0.12		
50	±0.05	500	±0.20		
100	±0.08	1000	±0.30		
		2000	±0.50		

[†] Tolerance values are from the ASTM E288, E542, and E694 standards.

Table 4.3   Measurement Errors for Digital Pipets[†]

Pipet Range	Volume (mL or μL)[‡]	Percent Measurement Error
10–100 μL	10	±3.0%
	50	±1.0%
	100	±0.8%
100–1000 μL	100	±3.0%
	500	±1.0%
	1000	±0.6%
1–10 mL	1	±3.0%
	5	±0.8%
	10	±0.6%

[†] Values are from www.eppendorf.com. [‡] Units for volume match the units for the pipet's range.

Table 4.4	Measurement Errors for Selected Balances	
Balance	Capacity (g)	Measurement Error
Precisa 160M	160	±1 mg
A & D ER 120M	120	±0.1 mg
Metler H54	160	±0.01 mg

PERSONAL ERRORS

Finally, analytical work is always subject to PERSONAL ERROR, examples of which include the ability to see a change in the color of an indicator that signals the endpoint of a titration, biases, such as consistently overestimating or underestimating the value on an instrument's readout scale, failing to calibrate instrumentation, and misinterpreting procedural directions. You can minimize personal errors by taking proper care.

IDENTIFYING DETERMINATE ERRORS

Determinate errors often are difficult to detect. Without knowing the expected value for an analysis, the usual situation in any analysis that matters, we often have nothing to which we can compare our experimental result. Nevertheless, there are strategies we can use to detect determinate errors.

The magnitude of a CONSTANT DETERMINATE ERROR is the same for all samples and is more significant when we analyze smaller samples. Analyzing samples of different sizes, therefore, allows us to detect a constant determinate error. For example, consider a quantitative analysis in which we separate the analyte from its matrix and determine its mass. Let's assume the sample is 50.0% w/w analyte. As we see in Table 4.5, the expected amount of analyte in a 0.100 g sample is 0.050 g. If the analysis has a positive constant determinate error of 0.010 g, then analyzing the sample gives 0.060 g of analyte, or a concentration of 60.0% w/w. As we increase the size of the sample the experimental results become closer to the expected result. An upward or downward trend in a graph of the analyte's experi-

| Table 4.5	Effect of a Constant Determinate Error on the Analysis of a Sample That is 50.0% w/w Analyte			
Mass Sample (g)	Expected Mass of Analyte (g)	Constant Error (g)	Experimental Mass of Analyte (g)	Experimental Concentration of Analyte (%w/w)
0.100	0.050	0.010	0.060	60.0
0.200	0.100	0.010	0.110	55.0
0.400	0.200	0.010	0.210	52.5
0.800	0.400	0.010	0.410	51.2
1.600	0.800	0.010	0.810	50.6
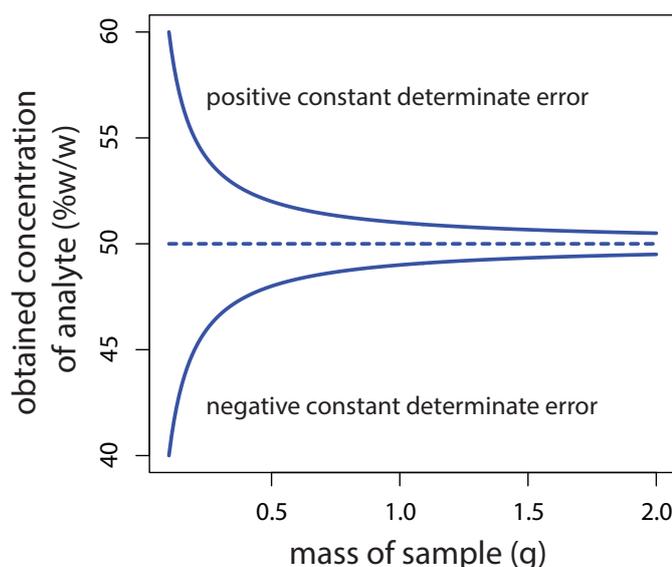
Figure 4.3 Effect of a constant positive determinate error of +0.01 g and a constant negative determinate error of –0.01 g on the determination of an analyte in samples of varying size. The analyte's expected concentration of 50% w/w is shown by the dashed line.

mental concentration versus the sample's mass (Figure 4.3) is evidence of a constant determinate error.

A PROPORTIONAL DETERMINATE ERROR, in which the error's magnitude depends on the amount of sample, is more difficult to detect because the result of the analysis is independent of the amount of sample. Table 4.6 outlines an example that shows the effect of a positive proportional error of 1.0% on the analysis of a sample that is 50.0% w/w in analyte. Regardless of the sample's size, each analysis gives the same result of 50.5% w/w analyte.

One approach for detecting a proportional determinate error is to analyze a standard that contains a known amount of analyte in a matrix similar to our samples. Standards are available from a variety of sources, such as the National Institute of Standards and Technology (where they are called STANDARD REFERENCE MATERIALS) or the American Society for Testing and Materials. Table 4.7, for example, lists certified values for several analytes in a standard sample of *Gingko biloba* leaves. Another approach is to compare our analysis to an analysis carried out using an independent analytical method that is known to give accurate results. If the two methods give significantly different results, then a determinate error is the likely cause.

Table 4.6	Effect of a Proportional Determinate Error on the Analysis of a Sample That is 50.0% w/w Analyte			
Mass Sample (g)	Expected Mass of Analyte (g)	Proportional Error (%)	Experimental Mass of Analyte (g)	Experimental Concentration of Analyte (%w/w)
0.100	0.050	1.00	0.0505	50.5
0.200	0.100	1.00	0.101	50.5
0.400	0.200	1.00	0.202	50.5
0.800	0.400	1.00	0.404	50.5
1.600	0.800	1.00	0.808	50.5

Table 4.7 Certified Concentrations for SRM 3246: *Ginkgo biloba* (Leaves)[†]				
Class of Analyte	Analyte	Mass Fraction (mg/g or ng/g)		
Flavonoids/Ginkgolide B	Quercetin	2.69	±	0.31
(mass fractions in mg/g)	Kaempferol	3.02	±	0.41
	Isorhamnetin	0.517	±	0.099
	Total Aglycones	6.22	±	0.77
Selected Terpenes	Ginkgolide A	0.57	±	0.28
(mass fractions in mg/g)	Ginkgolide B	0.470	±	0.090
	Ginkgolide C	0.59	±	0.22
	Ginkgolide J	0.18	±	0.10
	Biloabalide	1.52	±	0.40
	Total Terpene Lactones	3.3	±	1.1
Selected Toxic Elements	Cadmium	20.8	±	1.0
(mass fractions in ng/g)	Lead	995	±	30
	Mercury	23.08	±	0.17

[†] The primary purpose of this Standard Reference Material is to validate analytical methods for determining flavonoids, terpene lactones, and toxic elements in *Ginkgo biloba* or other materials with a similar matrix. Values are from the official Certificate of Analysis available at www.nist.gov.

Constant and proportional determinate errors have distinctly different sources, which we can define in terms of the relationship between the signal and the moles or concentration of analyte (equation 4.4 and equation 4.5). An invalid method blank, S_{mb}, is a constant determinate error as it adds or subtracts the same value to the signal. A poorly calibrated method, which yields an invalid sensitivity for the analyte, k_A, results in a proportional determinate error.

4B.2   Errors That Affect Precision

As we saw in Section 4A.2, precision is a measure of the spread of individual measurements or results about a central value, which we express as a range, a standard deviation, or a variance. Here we draw a distinction between two types of precision: repeatability and reproducibility. Repeatability is the precision when a single analyst completes an analysis in a single session using the same solutions, equipment, and instrumentation. Reproducibility, on the other hand, is the precision under any other set of conditions, including between analysts or between laboratory sessions for a single analyst. Since reproducibility includes additional sources of variability, the reproducibility of an analysis cannot be better than its repeatability.

Errors that affect precision are indeterminate and are characterized by random variations in their magnitude and their direction. Because they are random, positive and negative indeterminate errors tend to cancel, provided that we make a sufficient number of measurements. In such situ-

The ratio of the standard deviation associated with reproducibility to the standard deviation associated with repeatability is called the Horowitz ratio. For a wide variety of analytes in foods, for example, the median Horowtiz ratio is 2.0 with larger values for fatty acids and for trace elements; see Thompson, M.; Wood, R. "The 'Horowitz Ratio'–A Study of the Ratio Between Reproducibility and Repeatability in the Analysis of Foodstuffs," *Anal. Methods*, **2015**, *7*, 375–379.

ations the mean and the median largely are unaffected by the precision of the analysis.

SOURCES OF INDETERMINATE ERROR

We can assign indeterminate errors to several sources, including collecting samples, manipulating samples during the analysis, and making measurements. When we collect a sample, for instance, only a small portion of the available material is taken, which increases the chance that small-scale inhomogeneities in the sample will affect repeatability. Individual pennies, for example, may show variations in mass from several sources, including the manufacturing process and the loss of small amounts of metal or the addition of dirt during circulation. These variations are sources of indeterminate sampling errors.

During an analysis there are many opportunities to introduce indeterminate method errors. If our method for determining the mass of a penny includes directions for cleaning them of dirt, then we must be careful to treat each penny in the same way. Cleaning some pennies more vigorously than others might introduce an indeterminate method error.

Finally, all measuring devices are subject to indeterminate measurement errors due to limitations in our ability to read its scale. For example, a buret with scale divisions every 0.1 mL has an inherent indeterminate error of ±0.01–0.03 mL when we estimate the volume to the hundredth of a milliliter (Figure 4.4).

Figure 4.4 Close-up of a buret showing the difficulty in estimating volume. With scale divisions every 0.1 mL it is difficult to read the actual volume to better than ±0.01–0.03 mL.

EVALUATING INDETERMINATE ERROR

Indeterminate errors associated with our analytical equipment or instrumentation generally are easy to estimate if we measure the standard deviation for several replicate measurements, or if we monitor the signal's fluctuations over time in the absence of analyte (Figure 4.5) and calculate the standard deviation. Other sources of indeterminate error, such as treating samples inconsistently, are more difficult to estimate.
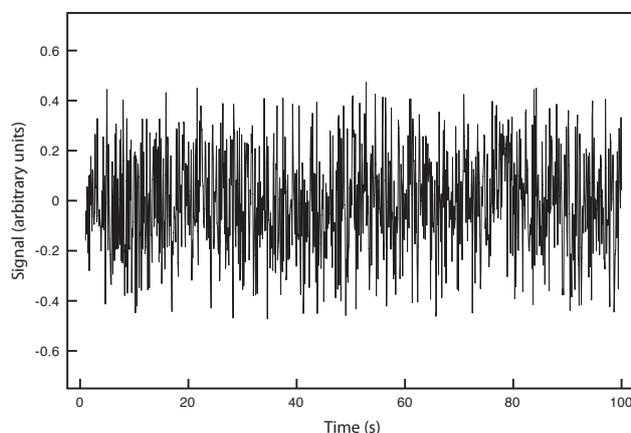
Figure 4.5 Background noise in an instrument showing the random fluctuations in the signal.

Table 4.8	Replicate Determinations of the Mass of a Single Circulating U. S. Penny		
Replicate	Mass (g)	Replicate	Mass (g)
1	3.025	6	3.023
2	3.024	7	3.022
3	3.028	8	3.021
4	3.027	9	3.026
5	3.028	10	3.024

To evaluate the effect of an indeterminate measurement error on our analysis of the mass of a circulating United States penny, we might make several determinations of the mass for a single penny (Table 4.8). The standard deviation for our original experiment (see Table 4.1) is 0.051 g, and it is 0.0024 g for the data in Table 4.8. The significantly better precision when we determine the mass of a single penny suggests that the precision of our analysis is not limited by the balance. A more likely source of indeterminate error is a variability in the masses of individual pennies.

In Section 4E we will discuss a statistical method—the *F*-test—that you can use to show that this difference is significant.

4B.3   Error and Uncertainty

Analytical chemists make a distinction between error and uncertainty.[3] ERROR is the difference between a single measurement or result and its expected value. In other words, error is a measure of BIAS. As discussed earlier, we divide errors into determinate and indeterminate sources. Although we can find and correct a source of determinate error, the indeterminate portion of the error remains.

UNCERTAINTY expresses the range of possible values for a measurement or result. Note that this definition of uncertainty is not the same as our definition of precision. We calculate precision from our experimental data and use it to estimate the magnitude of indeterminate errors. Uncertainty accounts for all errors—both determinate and indeterminate—that reasonably might affect a measurement or a result. Although we always try to correct determinate errors before we begin an analysis, the correction itself is subject to uncertainty.

Here is an example to help illustrate the difference between precision and uncertainty. Suppose you purchase a 10-mL Class A pipet from a laboratory supply company and use it without any additional calibration. The pipet's tolerance of ±0.02 mL is its uncertainty because your best estimate of its expected volume is 10.00 mL ± 0.02 mL. This uncertainty primarily is determinate. If you use the pipet to dispense several replicate samples of a solution and determine the volume of each sample, the resulting standard deviation is the pipet's precision. Table 4.9 shows results for ten such trials, with a mean of 9.992 mL and a standard deviation of ±0.006 mL. This standard deviation is the precision with which we expect to deliver a solu-

See Table 4.2 for the tolerance of a 10-mL class A transfer pipet.

3   Ellison, S.; Wegscheider, W.; Williams, A. *Anal. Chem.* **1997**, *69*, 607A–613A.

Table 4.9 Experimental Results for Volume Delivered by a 10-mL Class A Transfer Pipet			
Number	Volume (mL)	Number	Volume (mL)
1	10.002	6	9.983
2	9.993	7	9.991
3	9.984	8	9.990
4	9.996	9	9.988
5	9.989	10	9.999

tion using a Class A 10-mL pipet. In this case the pipet's published uncertainty of ±0.02 mL is worse than its experimentally determined precision of ±0.006 ml. Interestingly, the data in Table 4.9 allows us to calibrate this specific pipet's delivery volume as 9.992 mL. If we use this volume as a better estimate of the pipet's expected volume, then its uncertainty is ±0.006 mL. As expected, calibrating the pipet allows us to decrease its uncertainty.[4]

4C   Propagation of Uncertainty

Suppose we dispense 20 mL of a reagent using the Class A 10-mL pipet whose calibration information is given in Table 4.9. If the volume and uncertainty for one use of the pipet is 9.992 ± 0.006 mL, what is the volume and uncertainty if we use the pipet twice?

As a first guess, we might simply add together the volume and the maximum uncertainty for each delivery; thus

$$(9.992 \text{ mL} + 9.992 \text{ mL}) \pm (0.006 \text{ mL} + 0.006 \text{ mL}) = 19.984 \pm 0.012 \text{ mL}$$

It is easy to appreciate that combining uncertainties in this way overestimates the total uncertainty. Adding the uncertainty for the first delivery to that of the second delivery assumes that with each use the indeterminate error is in the same direction and is as large as possible. At the other extreme, we might assume that the uncertainty for one delivery is positive and the other is negative. If we subtract the maximum uncertainties for each delivery,

$$(9.992 \text{ mL} + 9.992 \text{ mL}) \pm (0.006 \text{ mL} - 0.006 \text{ mL}) = 19.984 \pm 0.000 \text{ mL}$$

we clearly underestimate the total uncertainty.

So what is the total uncertainty? From the discussion above, we reasonably expect that the total uncertainty is greater than ±0.000 mL and that it is less than ±0.012 mL. To estimate the uncertainty we use a mathematical technique known as the propagation of uncertainty. Our treatment of the propagation of uncertainty is based on a few simple rules.

Although we will not derive or further justify the rules presented in this section, you may consult this chapter's additional resources for references that discuss the propagation of uncertainty in more detail.

4   Kadis, R. *Talanta* **2004**, *64*, 167–173.

4C.1   A Few Symbols

A PROPAGATION OF UNCERTAINTY allows us to estimate the uncertainty in a result from the uncertainties in the measurements used to calculate that result. For the equations in this section we represent the result with the symbol R, and we represent the measurements with the symbols A, B, and C. The corresponding uncertainties are u_R, u_A, u_B, and u_C. We can define the uncertainties for A, B, and C using standard deviations, ranges, or tolerances (or any other measure of uncertainty), as long as we use the same form for all measurements.

The requirement that we express each uncertainty in the same way is a critically important point. Suppose you have a range for one measurement, such as a pipet's tolerance, and standard deviations for the other measurements. All is not lost. There are ways to convert a range to an estimate of the standard deviation. See Appendix 2 for more details.

4C.2   Uncertainty When Adding or Subtracting

When we add or subtract measurements we propagate their absolute uncertainties. For example, if the result is given by the equation

$$R = A + B - C$$

then the absolute uncertainty in R is

$$u_R = \sqrt{u_A^2 + u_B^2 + u_C^2} \qquad\qquad 4.6$$

Example 4.5

If we dispense 20 mL using a 10-mL Class A pipet, what is the total volume dispensed and what is the uncertainty in this volume? First, complete the calculation using the manufacturer's tolerance of $10.00 \text{ mL} \pm 0.02 \text{ mL}$, and then using the calibration data from Table 4.9.

Solution

To calculate the total volume we add the volumes for each use of the pipet. When using the manufacturer's values, the total volume is

$$V = 10.00 \text{ mL} + 10.00 \text{ mL} = 20.00 \text{ mL}$$

and when using the calibration data, the total volume is

$$V = 9.992 \text{ mL} + 9.992 \text{ mL} = 19.984 \text{ mL}$$

Using the pipet's tolerance as an estimate of its uncertainty gives the uncertainty in the total volume as

$$u_R = \sqrt{(0.02)^2 + (0.02)^2} = 0.028 \text{ mL}$$

and using the standard deviation for the data in Table 4.9 gives an uncertainty of

$$u_R = \sqrt{(0.006)^2 + (0.006)^2} = 0.0085 \text{ mL}$$

Rounding the volumes to four significant figures gives $20.00 \text{ mL} \pm 0.03 \text{ mL}$ when we use the tolerance values, and $19.98 \pm 0.01 \text{ mL}$ when we use the calibration data.

4C.3   Uncertainty When Multiplying or Dividing

When we multiple or divide measurements we propagate their relative uncertainties. For example, if the result is given by the equation

$$R = \frac{A \times B}{C}$$

then the relative uncertainty in R is

$$\frac{u_R}{R} = \sqrt{\left(\frac{u_A}{A}\right)^2 + \left(\frac{u_B}{B}\right)^2 + \left(\frac{u_C}{C}\right)^2} \qquad 4.7$$

Example 4.6

The quantity of charge, Q, in coulombs that passes through an electrical circuit is

$$Q = i \times t$$

where i is the current in amperes and t is the time in seconds. When a current of $0.15\ \text{A} \pm 0.01\ \text{A}$ passes through the circuit for $120\ \text{s} \pm 1\ \text{s}$, what is the total charge and its uncertainty?

SOLUTION

The total charge is

$$Q = (0.15\ \text{A}) \times (120\ \text{s}) = 18\ \text{C}$$

Since charge is the product of current and time, the relative uncertainty in the charge is

$$\frac{u_R}{R} = \sqrt{\left(\frac{0.01}{0.15}\right)^2 + \left(\frac{1}{120}\right)^2} = 0.0672$$

and the charge's absolute uncertainty is

$$u_R = R \times 0.0672 = (18\ \text{C}) \times (0.0672) = 1.2\ \text{C}$$

Thus, we report the total charge as $18\ \text{C} \pm 1\ \text{C}$.

4C.4   Uncertainty for Mixed Operations

Many chemical calculations involve a combination of adding and subtracting, and of multiply and dividing. As shown in the following example, we can calculate the uncertainty by separately treating each operation using equation 4.6 and equation 4.7 as needed.

Example 4.7

For a concentration technique, the relationship between the signal and the an analyte's concentration is

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

What is the analyte's concentration, C_A, and its uncertainty if S_{total} is 24.37 ± 0.02, S_{mb} is 0.96 ± 0.02, and k_A is 0.186 ± 0.003 ppm^{-1}?

SOLUTION

Rearranging the equation and solving for C_A

$$C_A = \frac{S_{total} - S_{mb}}{k_A} = \frac{24.37 - 0.96}{0.186 \text{ ppm}^{-1}} = \frac{23.41}{0.186 \text{ ppm}^{-1}} = 125.9 \text{ ppm}$$

gives the analyte's concentration as 126 ppm. To estimate the uncertainty in C_A, we first use equation 4.6 to determine the uncertainty for the numerator.

$$u_R = \sqrt{(0.02)^2 + (0.02)^2} = 0.028$$

The numerator, therefore, is 23.41 ± 0.028. To complete the calculation we use equation 4.7 to estimate the relative uncertainty in C_A.

$$\frac{u_R}{R} = \sqrt{\left(\frac{0.028}{23.41}\right)^2 + \left(\frac{0.003}{0.186}\right)^2} = 0.0162$$

The absolute uncertainty in the analyte's concentration is

$$u_R = (125.9 \text{ ppm}) \times (0.0162) = 2.0 \text{ ppm}$$

Thus, we report the analyte's concentration as 126 ppm ± 2 ppm.

4C.5 Uncertainty for Other Mathematical Functions

Many other mathematical operations are common in analytical chemistry, including the use of powers, roots, and logarithms. Table 4.10 provides equations for propagating uncertainty for some of these function.

Example 4.8

If the pH of a solution is 3.72 with an absolute uncertainty of ± 0.03, what is the $[H^+]$ and its uncertainty?

SOLUTION

The concentration of H^+ is

$$[H^+] = 10^{-pH} = 10^{-3.72} = 1.91 \times 10^{-4} \text{ M}$$

Practice Exercise 4.2

To prepare a standard solution of Cu^{2+} you obtain a piece of copper from a spool of wire. The spool's initial weight is 74.2991 g and its final weight is 73.3216 g. You place the sample of wire in a 500 mL volumetric flask, dissolve it in 10 mL of HNO_3, and dilute to volume. Next, you pipet a 1 mL portion to a 250-mL volumetric flask and dilute to volume. What is the final concentration of Cu^{2+} in mg/L, and its uncertainty? Assume that the uncertainty in the balance is ± 0.1 mg and that you are using Class A glassware.

Click here when to review your answer to this exercise.

Table 4.10 Propagation of Uncertainty for Selected Mathematical Functions[†]	
Function	u_R
$R = kA$	$u_R = ku_A$
$R = A + B$	$u_R = \sqrt{u_A^2 + u_B^2}$
$R = A - B$	$u_R = \sqrt{u_A^2 + u_B^2}$
$R = A \times B$	$\dfrac{u_R}{R} = \sqrt{\left(\dfrac{u_A}{A}\right)^2 + \left(\dfrac{u_B}{B}\right)^2}$
$R = \dfrac{A}{B}$	$\dfrac{u_R}{R} = \sqrt{\left(\dfrac{u_A}{A}\right)^2 + \left(\dfrac{u_B}{B}\right)^2}$
$R = \ln(A)$	$u_R = \dfrac{u_A}{A}$
$R = \log(A)$	$u_R = 0.4343 \times \dfrac{u_A}{A}$
$R = e^A$	$\dfrac{u_R}{R} = u_A$
$R = 10^A$	$\dfrac{u_R}{R} = 2.303 \times u_A$
$R = A^k$	$\dfrac{u_R}{R} = k \times \dfrac{u_A}{A}$

[†] Assumes that the measurements A and B are independent; k is a constant whose value has no uncertainty.

or 1.9×10^{-4} M to two significant figures. From Table 4.10 the relative uncertainty in $[H^+]$ is

$$\frac{u_R}{R} = 2.303 \times u_A = 2.303 \times 0.03 = 0.069$$

The uncertainty in the concentration, therefore, is

$$(1.91 \times 10^{-4} \text{ M}) \times (0.069) = 1.3 \times 10^{-5} \text{ M}$$

We report the $[H^+]$ as $1.9\ (\pm 0.1) \times 10^{-4}$ M.

Writing this result as
$$1.9\ (\pm 0.1) \times 10^{-4} \text{ M}$$
is equivalent to
$$1.9 \times 10^{-4} \text{ M} \pm 0.1 \times 10^{-4} \text{ M}$$

Practice Exercise 4.3

A solution of copper ions is blue because it absorbs yellow and orange light. Absorbance, A, is defined as

$$A = -\log\left(\frac{P}{P_o}\right)$$

where P_o is the power of radiation as emitted from the light source and P is its power after it passes through the solution. What is the absorbance if P_o is 3.80×10^2 and P is 1.50×10^2? If the uncertainty in measuring P_o and P is 15, what is the uncertainty in the absorbance?

Click here to review your answer to this exercise.

4C.6 Is Calculating Uncertainty Actually Useful?

Given the effort it takes to calculate uncertainty, it is worth asking whether such calculations are useful. The short answer is, yes. Let's consider three examples of how we can use a propagation of uncertainty to help guide the development of an analytical method.

One reason to complete a propagation of uncertainty is that we can compare our estimate of the uncertainty to that obtained experimentally. For example, to determine the mass of a penny we measure its mass twice— once to tare the balance at 0.000 g and once to measure the penny's mass. If the uncertainty in each measurement of mass is ± 0.001 g, then we estimate the total uncertainty in the penny's mass as

$$u_R = \sqrt{(0.001)^2 + (0.001)^2} = 0.0014 \text{ g}$$

If we measure a single penny's mass several times and obtain a standard deviation of ± 0.050 g, then we have evidence that the measurement process is out of control. Knowing this, we can identify and correct the problem.

We also can use a propagation of uncertainty to help us decide how to improve an analytical method's uncertainty. In Example 4.7, for instance, we calculated an analyte's concentration as 126 ppm ± 2 ppm, which is a percent uncertainty of 1.6%. Suppose we want to decrease the percent uncertainty to no more than 0.8%. How might we accomplish this? Looking back at the calculation, we see that the concentration's relative uncertainty is determined by the relative uncertainty in the measured signal (corrected for the reagent blank)

$$\frac{0.028}{23.41} = 0.0012 \text{ or } 0.12\%$$

and the relative uncertainty in the method's sensitivity, k_A,

$$\frac{0.003 \text{ ppm}^{-1}}{0.186 \text{ ppm}^{-1}} = 0.016 \text{ or } 1.6\%$$

Of these two terms, the uncertainty in the method's sensitivity dominates the overall uncertainty. Improving the signal's uncertainty will not improve the overall uncertainty of the analysis. To achieve an overall uncertainty of 0.8% we must improve the uncertainty in k_A to ± 0.0015 ppm^{-1}.

Practice Exercise 4.4

Verify that an uncertainty of ± 0.0015 ppm^{-1} for k_A is the correct result.

Click here to review your answer to this exercise.

Finally, we can use a propagation of uncertainty to determine which of several procedures provides the smallest uncertainty. When we dilute a stock solution usually there are several combinations of volumetric glassware that will give the same final concentration. For instance, we can dilute a stock solution by a factor of 10 using a 10-mL pipet and a 100-mL volumetric

flask, or using a 25-mL pipet and a 250-mL volumetric flask. We also can accomplish the same dilution in two steps using a 50-mL pipet and 100-mL volumetric flask for the first dilution, and a 10-mL pipet and a 50-mL volumetric flask for the second dilution. The overall uncertainty in the final concentration—and, therefore, the best option for the dilution—depends on the uncertainty of the volumetric pipets and volumetric flasks. As shown in the following example, we can use the tolerance values for volumetric glassware to determine the optimum dilution strategy.[5]

Example 4.9

Which of the following methods for preparing a 0.0010 M solution from a 1.0 M stock solution provides the smallest overall uncertainty?

(a) A one-step dilution that uses a 1-mL pipet and a 1000-mL volumetric flask.

(b) A two-step dilution that uses a 20-mL pipet and a 1000-mL volumetric flask for the first dilution, and a 25-mL pipet and a 500-mL volumetric flask for the second dilution.

SOLUTION

The dilution calculations for case (a) and case (b) are

$$\text{case (a):} 1.0 \text{ M} \times \frac{1.000 \text{ mL}}{1000.0 \text{ mL}} = 0.0010 \text{ M}$$

$$\text{case (b):} 1.0 \text{ M} \times \frac{20.00 \text{ mL}}{1000.0 \text{ mL}} \times \frac{25.00 \text{ mL}}{500.0 \text{ mL}} = 0.0010 \text{ M}$$

Using tolerance values from Table 4.2, the relative uncertainty for case (a) is

$$\frac{u_R}{R} = \sqrt{\left(\frac{0.006}{1.000}\right)^2 + \left(\frac{0.3}{1000.0}\right)^2} = 0.006$$

and for case (b) the relative uncertainty is

$$\frac{u_R}{R} = \sqrt{\left(\frac{0.03}{20.00}\right)^2 + \left(\frac{0.3}{1000.0}\right)^2 + \left(\frac{0.03}{25.00}\right)^2 + \left(\frac{0.2}{500.0}\right)^2} = 0.002$$

Since the relative uncertainty for case (b) is less than that for case (a), the two-step dilution provides the smallest overall uncertainty.

Of course we must balance the smaller uncertainty for case (b) against the increased opportunity for introducing a determinate error when making two dilutions instead of just one dilution, as in case (a).

4D   The Distribution of Measurements and Results

Earlier we reported results for a determination of the mass of a circulating United States penny, obtaining a mean of 3.117 g and a standard deviation of 0.051 g. Table 4.11 shows results for a second, independent determination of a penny's mass, as well as the data from the first experiment. Although the means and standard deviations for the two experiments are similar, they are not identical. The difference between the two experiments

5   Lam, R. B.; Isenhour, T. L. *Anal. Chem.* **1980**, *52*, 1158–1161.

Table 4.11 Results for Two Determinations of the Mass of a Circulating United States Penny			
First Experiment		Second Experiment	
Penny	Mass (g)	Penny	Mass (g)
1	3.080	1	3.052
2	3.094	2	3.141
3	3.107	3	3.083
4	3.056	4	3.083
5	3.112	5	3.048
6	3.174		
7	3.198		
$\overline{X}$	3.117		3.081
s	0.051		0.037

raises some interesting questions. Are the results for one experiment better than the results for the other experiment? Do the two experiments provide equivalent estimates for the mean and the standard deviation? What is our best estimate of a penny's expected mass? To answer these questions we need to understand how we might predict the properties of all pennies using the results from an analysis of a small sample of pennies. We begin by making a distinction between populations and samples.

4D.1 Populations and Samples

A POPULATION is the set of all objects in the system we are investigating. For the data in Table 4.11, the population is all United States pennies in circulation. This population is so large that we cannot analyze every member of the population. Instead, we select and analyze a limited subset, or SAMPLE of the population. The data in Table 4.11, for example, shows the results for two such samples drawn from the larger population of all circulating United States pennies.

4D.2 Probability Distributions for Populations

Table 4.11 provides the means and the standard deviations for two samples of circulating United States pennies. What do these samples tell us about the population of pennies? What is the largest possible mass for a penny? What is the smallest possible mass? Are all masses equally probable, or are some masses more common?

To answer these questions we need to know how the masses of individual pennies are distributed about the population's average mass. We represent the distribution of a population by plotting the probability or frequency of

obtaining a specific result as a function of the possible results. Such plots are called PROBABILITY DISTRIBUTIONS.

There are many possible probability distributions; in fact, the probability distribution can take any shape depending on the nature of the population. Fortunately many chemical systems display one of several common probability distributions. Two of these distributions, the binomial distribution and the normal distribution, are discussed in this section.

BINOMIAL DISTRIBUTION

The BINOMIAL DISTRIBUTION describes a population in which the result is the number of times a particular event occurs during a fixed number of trials. Mathematically, the binomial distribution is defined as

$$P(X,N) = \frac{N!}{X!(N-X)!} \times p^X \times (1-p)^{N-X}$$

where $P(X,N)$ is the probability that an event occurs X times during N trials, and p is the event's probability for a single trial. If you flip a coin five times, $P(2,5)$ is the probability the coin will turn up "heads" exactly twice.

A binomial distribution has well-defined measures of central tendency and spread. The expected mean value is

$$\mu = Np$$

and the expected spread is given by the variance

$$\sigma^2 = Np(1-p)$$

or the standard deviation.

$$\sigma = \sqrt{Np(1-p)}$$

The binomial distribution describes a population whose members have only specific, discrete values. When you roll a die, for example, the possible values are 1, 2, 3, 4, 5, or 6. A roll of 3.45 is not possible. As shown in Example 4.10, one example of a chemical system that obeys the binomial distribution is the probability of finding a particular isotope in a molecule.

The term $N!$ reads as N-factorial and is the product $N \times (N–1) \times (N–2) \times \cdots \times 1$. For example, $4!$ is $4 \times 3 \times 2 \times 1 = 24$. Your calculator probably has a key for calculating factorials.

Example 4.10

Carbon has two stable, non-radioactive isotopes, ^{12}C and ^{13}C, with relative isotopic abundances of, respectively, 98.89% and 1.11%.

(a) What are the mean and the standard deviation for the number of ^{13}C atoms in a molecule of cholesterol ($C_{27}H_{44}O$)?

(b) What is the probability that a molecule of cholesterol has no atoms of ^{13}C?

SOLUTION

The probability of finding an atom of ^{13}C in a molecule of cholesterol follows a binomial distribution, where X is the number of ^{13}C atoms, N

is the number of carbon atoms in a molecule of cholesterol, and p is the probability that an atom of carbon in ^{13}C.

(a) The mean number of ^{13}C atoms in a molecule of cholesterol is

$$\mu = Np = 27 \times 0.0111 = 0.300$$

with a standard deviation of

$$\sigma = \sqrt{Np(1-p)} = \sqrt{27 \times 0.0111 \times (1-0.0111)} = 0.544$$

(b) The probability of finding a molecule of cholesterol without an atom of ^{13}C is

$$P(0,27) = \frac{27!}{0!(27-0)!} \times (0.0111)^0 \times$$
$$(1-0.0111)^{27-0} = 0.740$$

There is a 74.0% probability that a molecule of cholesterol will not have an atom of ^{13}C, a result consistent with the observation that the mean number of ^{13}C atoms per molecule of cholesterol, 0.300, is less than one.

A portion of the binomial distribution for atoms of ^{13}C in cholesterol is shown in Figure 4.6. Note in particular that there is little probability of finding more than two atoms of ^{13}C in any molecule of cholesterol.

NORMAL DISTRIBUTION

A binomial distribution describes a population whose members have only certain discrete values. This is the case with the number of ^{13}C atoms in cholesterol. A molecule of cholesterol, for example, can have two ^{13}C atoms, but it can not have 2.5 atoms of ^{13}C. A population is continuous if its members may take on any value. The efficiency of extracting cholesterol from a
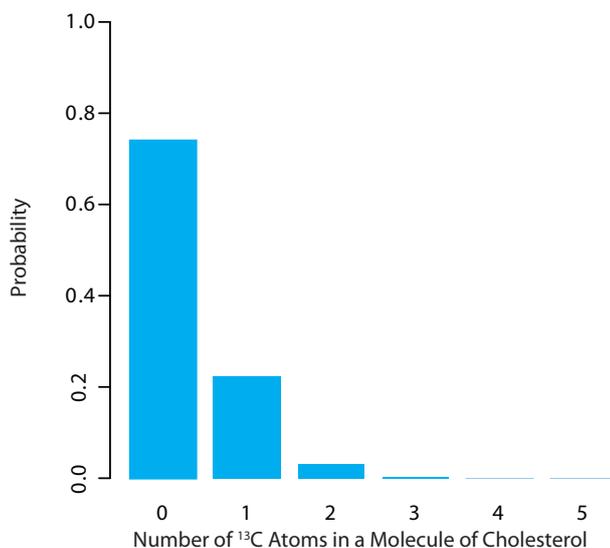
Figure 4.6 Portion of the binomial distribution for the number of naturally occurring ^{13}C atoms in a molecule of cholesterol. Only 3.6% of cholesterol molecules contain more than one atom of ^{13}C, and only 0.33% contain more than two atoms of ^{13}C.

sample, for example, can take on any value between 0% (no cholesterol is extracted) and 100% (all cholesterol is extracted).

The most common continuous distribution is the Gaussian, or NORMAL DISTRIBUTION, the equation for which is

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

where μ is the expected mean for a population with n members

$$\mu = \frac{\sum_{i=1}^{n} X_i}{n}$$

and σ^2 is the population's variance.

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n} \qquad 4.8$$

Examples of three normal distributions, each with an expected mean of 0 and with variances of 25, 100, or 400, respectively, are shown in Figure 4.7. Two features of these normal distribution curves deserve attention. First, note that each normal distribution has a single maximum that corresponds to μ, and that the distribution is symmetrical about this value. Second, increasing the population's variance increases the distribution's spread and decreases its height; the area under the curve, however, is the same for all three distributions.

The area under a normal distribution curve is an important and useful property as it is equal to the probability of finding a member of the population within a particular range of values. In Figure 4.7, for example, 99.99% of the population shown in curve (a) have values of X between –20 and +20. For curve (c), 68.26% of the population's members have values of X between –20 and +20.

Because a normal distribution depends solely on μ and σ^2, the probability of finding a member of the population between any two limits is

Figure 4.7 Normal distribution curves for:

(a) $\mu = 0$; $\sigma^2 = 25$
(b) $\mu = 0$; $\sigma^2 = 100$
(c) $\mu = 0$; $\sigma^2 = 400$
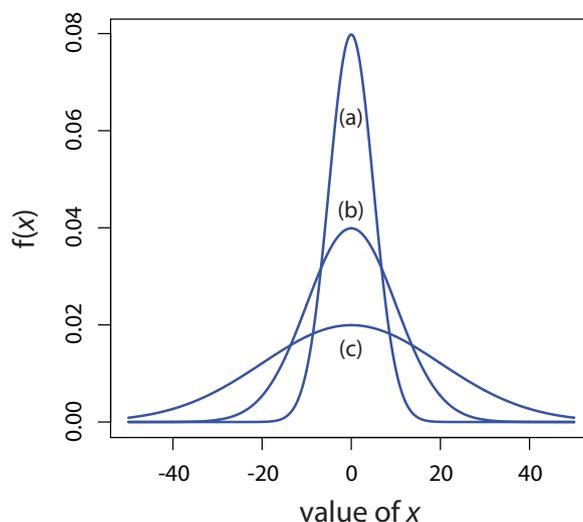
the same for all normally distributed populations. Figure 4.8, for example, shows that 68.26% of the members of a normal distribution have a value within the range $\mu \pm 1\sigma$, and that 95.44% of population's members have values within the range $\mu \pm 2\sigma$. Only 0.27% members of a population have values that exceed the expected mean by more than $\pm 3\sigma$. Additional ranges and probabilities are gathered together in the probability table included in Appendix 3. As shown in Example 4.11, if we know the mean and the standard deviation for a normally distributed population, then we can determine the percentage of the population between any defined limits.

Example 4.11

The amount of aspirin in the analgesic tablets from a particular manufacturer is known to follow a normal distribution with $\mu = 250$ mg and $\sigma = 5$. In a random sample of tablets from the production line, what percentage are expected to contain between 243 and 262 mg of aspirin?

SOLUTION

We do not determine directly the percentage of tablets between 243 mg and 262 mg of aspirin. Instead, we first find the percentage of tablets with less than 243 mg of aspirin and the percentage of tablets having more than 262 mg of aspirin. Subtracting these results from 100%, gives the percentage of tablets that contain between 243 mg and 262 mg of aspirin.
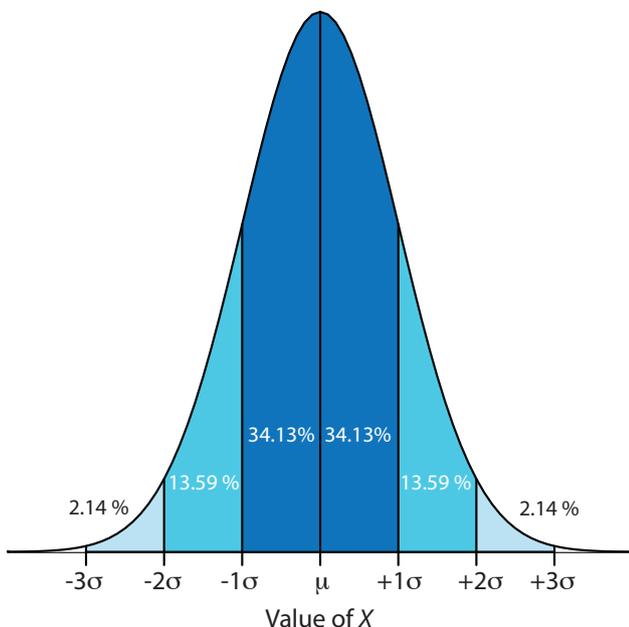
Figure 4.8 Normal distribution curve showing the area under the curve for several different ranges of values of X. As shown here, 68.26% of the members of a normally distributed population have values within $\pm 1\sigma$ of the population's expected mean, and 13.59% have values between $\mu - 1\sigma$ and $\mu - 2\sigma$. The area under the curve between any two limits is found using the probability table in Appendix 3.

To find the percentage of tablets with less than 243 mg of aspirin or more than 262 mg of aspirin we calculate the deviation, z, of each limit from μ in terms of the population's standard deviation, σ

$$z = \frac{X - \mu}{\sigma}$$

where X is the limit in question. The deviation for the lower limit is

$$z_{lower} = \frac{243 - 250}{5} = -1.4$$

and the deviation for the upper limit is

$$z_{upper} = \frac{262 - 250}{5} = +2.4$$

Using the table in Appendix 3, we find that the percentage of tablets with less than 243 mg of aspirin is 8.08%, and that the percentage of tablets with more than 262 mg of aspirin is 0.82%. Therefore, the percentage of tablets containing between 243 and 262 mg of aspirin is

$$100.00\% - 8.08\% - 0.82\% = 91.10\%$$

Figure 4.9 shows the distribution of aspiring in the tablets, with the area in blue showing the percentage of tablets containing between 243 mg and 262 mg of aspirin.

Practice Exercise 4.5

What percentage of aspirin tablets will contain between 240 mg and 245 mg of aspirin if the population's mean is 250 mg and the population's standard deviation is 5 mg.

Click here to review your answer to this exercise.
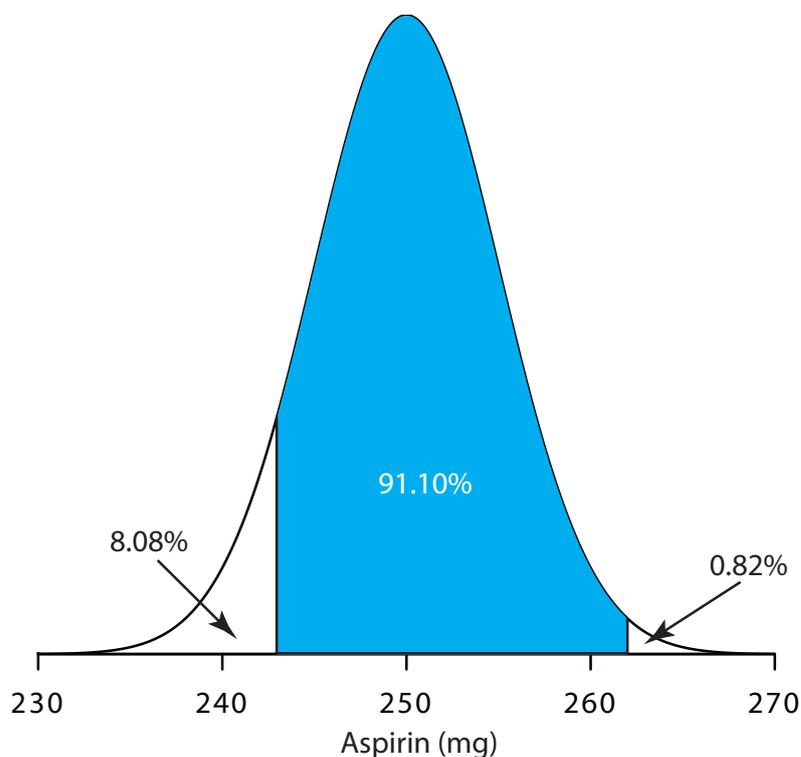
Figure 4.9 Normal distribution for the population of aspirin tablets in Example 4.11. The population's mean and standard deviation are 250 mg and 5 mg, respectively. The shaded area shows the percentage of tablets containing between 243 mg and 262 mg of aspirin.

4D.3 Confidence Intervals for Populations

If we select at random a single member from a population, what is its most likely value? This is an important question, and, in one form or another, it is at the heart of any analysis in which we wish to extrapolate from a sample to the sample's parent population. One of the most important features of a population's probability distribution is that it provides a way to answer this question.

Figure 4.8 shows that for a normal distribution, 68.26% of the population's members have values within the range $\mu \pm 1\sigma$. Stating this another way, there is a 68.26% probability that the result for a single sample drawn from a normally distributed population is in the interval $\mu \pm 1\sigma$. In general, if we select a single sample we expect its value, X_i is in the range

$$X_i = \mu \pm z\sigma \qquad 4.9$$

where the value of z is how confident we are in assigning this range. Values reported in this fashion are called CONFIDENCE INTERVALS. Equation 4.9, for example, is the confidence interval for a single member of a population. Table 4.12 gives the confidence intervals for several values of z. For reasons discussed later in the chapter, a 95% confidence level is a common choice in analytical chemistry.

When $z = 1$, we call this the 68.26% confidence interval.

Example 4.12

What is the 95% confidence interval for the amount of aspirin in a single analgesic tablet drawn from a population for which μ is 250 mg and for which σ is 5?

SOLUTION

Using Table 4.12, we find that z is 1.96 for a 95% confidence interval. Substituting this into equation 4.9 gives the confidence interval for a single tablet as

$$X_i = \mu \pm 1.96\sigma = 250 \text{ mg} \pm (1.96 \times 5) = 250 \text{ mg} \pm 10 \text{ mg}$$

Table 4.12 Confidence Intervals for a Normal Distribution ($\mu \pm z\sigma$)	
z	Confidence Interval (%)
0.50	38.30
1.00	68.26
1.50	86.64
1.96	95.00
2.00	95.44
2.50	98.76
3.00	99.73
3.50	99.95

A confidence interval of 250 mg $\pm$ 10 mg means that 95% of the tablets in the population contain between 240 and 260 mg of aspirin.

Alternatively, we can rewrite equation 4.9 so that it gives the confidence interval is for μ based on the population's standard deviation and the value of a single member drawn from the population.

$$\mu = X_i \pm z\sigma \qquad \qquad 4.10$$

Example 4.13

The population standard deviation for the amount of aspirin in a batch of analgesic tablets is known to be 7 mg of aspirin. If you randomly select and analyze a single tablet and find that it contains 245 mg of aspirin, what is the 95% confidence interval for the population's mean?

SOLUTION

The 95% confidence interval for the population mean is given as

$$\mu = X_i \pm z\sigma = 245 \text{ mg} \pm (1.96 \times 7) \text{ mg} = 245 \text{ mg} \pm 14 \text{ mg}$$

Therefore, based on this one sample, we estimate that there is 95% probability that the population's mean, μ, lies within the range of 231 mg to 259 mg of aspirin.

It is unusual to predict the population's expected mean from the analysis of a single sample; instead, we collect n samples drawn from a population of known σ, and report the mean, $\overline{X}$. The standard deviation of the mean, $\sigma_{\overline{X}}$, which also is known as the STANDARD ERROR OF THE MEAN, is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

The confidence interval for the population's mean, therefore, is

$$\mu = \overline{X} \pm \frac{z\sigma}{\sqrt{n}} \qquad \qquad 4.11$$

Example 4.14

What is the 95% confidence interval for the analgesic tablets in Example 4.13, if an analysis of five tablets yields a mean of 245 mg of aspirin?

SOLUTION

In this case the confidence interval is

$$\mu = 245 \text{ mg} \pm \frac{1.96 \times 7}{\sqrt{5}} \text{ mg} = 245 \text{ mg} \pm 6 \text{ mg}$$

We estimate a 95% probability that the population's mean is between 239 mg and 251 mg of aspirin. As expected, the confidence interval when using the mean of five samples is smaller than that for a single sample.

Note the qualification that the prediction for μ is based on one sample; a different sample likely will give a different 95% confidence interval. Our result here, therefore, is an estimate for μ based on this one sample.

Problem 8 at the end of the chapter asks you to derive this equation using a propagation of uncertainty.

Practice Exercise 4.6

An analysis of seven aspirin tablets from a population known to have a standard deviation of 5, gives the following results in mg aspirin per tablet:

246   249   255   251   251   247   250

What is the 95% confidence interval for the population's expected mean?

Click here when you are ready to review your answer.

4D.4   Probability Distributions for Samples

In Examples 4.11–4.14 we assumed that the amount of aspirin in analgesic tablets is normally distributed. Without analyzing every member of the population, how can we justify this assumption? In a situation where we cannot study the whole population, or when we cannot predict the mathematical form of a population's probability distribution, we must deduce the distribution from a limited sampling of its members.

SAMPLE DISTRIBUTIONS AND THE CENTRAL LIMIT THEOREM

Let's return to the problem of determining a penny's mass to explore further the relationship between a population's distribution and the distribution of a sample drawn from that population. The two sets of data in Table 4.11 are too small to provide a useful picture of a sample's distribution, so we will use the larger sample of 100 pennies shown in Table 4.13. The mean and the standard deviation for this sample are 3.095 g and 0.0346 g, respectively.

A HISTOGRAM (Figure 4.10) is a useful way to examine the data in Table 4.13. To create the histogram, we divide the sample into intervals, by mass, and determine the percentage of pennies within each interval (Table 4.14). Note that the sample's mean is the midpoint of the histogram.

Figure 4.10 also includes a normal distribution curve for the population of pennies, based on the assumption that the mean and the variance for the sample are appropriate estimates for the population's mean and variance. Although the histogram is not perfectly symmetric in shape, it provides a good approximation of the normal distribution curve, suggesting that the sample of 100 pennies is normally distributed. It is easy to imagine that the histogram will approximate more closely a normal distribution if we include additional pennies in our sample.

We will not offer a formal proof that the sample of pennies in Table 4.13 and the population of all circulating U. S. pennies are normally distributed; however, the evidence in Figure 4.10 strongly suggests this is true. Although we cannot claim that the results of all experiments are normally distributed, in most cases our data are normally distributed. According to the CENTRAL LIMIT THEOREM, when a measurement is subject to a variety of indeterminate errors, the results for that measurement will approximate

Table 4.13   Masses for a Sample of 100 Circulating U. S. Pennies

Penny	Mass (g)	Penny	Mass (g)	Penny	Mass (g)	Penny	Mass (g)
1	3.126	26	3.073	51	3.101	76	3.086
2	3.140	27	3.084	52	3.049	77	3.123
3	3.092	28	3.148	53	3.082	78	3.115
4	3.095	29	3.047	54	3.142	79	3.055
5	3.080	30	3.121	55	3.082	80	3.057
6	3.065	31	3.116	56	3.066	81	3.097
7	3.117	32	3.005	57	3.128	82	3.066
8	3.034	33	3.115	58	3.112	83	3.113
9	3.126	34	3.103	59	3.085	84	3.102
10	3.057	35	3.086	60	3.086	85	3.033
11	3.053	36	3.103	61	3.084	86	3.112
12	3.099	37	3.049	62	3.104	87	3.103
13	3.065	38	2.998	63	3.107	88	3.198
14	3.059	39	3.063	64	3.093	89	3.103
15	3.068	40	3.055	65	3.126	90	3.126
16	3.060	41	3.181	66	3.138	91	3.111
17	3.078	42	3.108	67	3.131	92	3.126
18	3.125	43	3.114	68	3.120	93	3.052
19	3.090	44	3.121	69	3.100	94	3.113
20	3.100	45	3.105	70	3.099	95	3.085
21	3.055	46	3.078	71	3.097	96	3.117
22	3.105	47	3.147	72	3.091	97	3.142
23	3.063	48	3.104	73	3.077	98	3.031
24	3.083	49	3.146	74	3.178	99	3.083
25	3.065	50	3.095	75	3.054	100	3.104

Table 4.14   Frequency Distribution for the Data in Table 4.13

Mass Interval	Frequency (as %)	Mass Interval	Frequency (as %)
2.991–3.009	2	3.105–3.123	19
3.010–3.028	0	3.124–3.142	12
3.029–3.047	4	3.143–3.161	3
3.048–3.066	19	3.162–3.180	1
3.067–3.085	14	3.181–3.199	2
3.086–3.104	24		
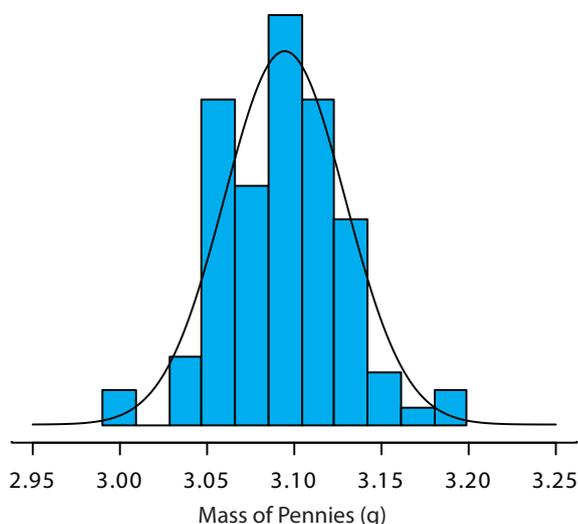
Figure 4.10 The blue bars show a histogram for the data in Table 4.13. The height of each bar corresponds to the percentage of pennies within one of the mass intervals in Table 4.14. Superimposed on the histogram is a normal distribution curve based on the assumption that μ and σ^2 for the population are equivalent to $\overline{X}$ and s^2 for the sample. The total area of the histogram's bars and the area under the normal distribution curve are equal.

a normal distribution.[6] The central limit theorem holds true even if the individual sources of indeterminate error are not normally distributed. The chief limitation to the central limit theorem is that the sources of indeterminate error must be independent and of similar magnitude so that no one source of error dominates the final distribution.

An additional feature of the central limit theorem is that a distribution of means for samples drawn from a population with any distribution will approximate closely a normal distribution if the size of each sample is sufficiently large. For example, Figure 4.11 shows the distribution for two samples of 10 000 drawn from a uniform distribution in which every value between 0 and 1 occurs with an equal frequency. For samples of size $n = 1$, the resulting distribution closely approximates the population's uniform distribution. The distribution of the means for samples of size $n = 10$, however, closely approximates a normal distribution.

DEGREES OF FREEDOM

Did you notice the differences between the equation for the variance of a population and the variance of a sample? If not, here are the two equations:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}$$

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n - 1}$$

Both equations measure the variance around the mean, using μ for a population and $\overline{X}$ for a sample. Although the equations use different measures for the mean, the intention is the same for both the sample and the popu-

You might reasonably ask whether this aspect of the central limit theorem is important as it is unlikely that we will complete 10 000 analyses, each of which is the average of 10 individual trials. This is deceiving. When we acquire a sample of soil, for example, it consists of many individual particles each of which is an individual sample of the soil. Our analysis of this sample, therefore, gives the mean for this large number of individual soil particles. Because of this, the central limit theorem is relevant.

For a discussion of circumstances where the central limit theorem may not apply, see "Do You Reckon It's Normally Distributed?", the full reference for which is Majewsky, M.; Wagner, M.; Farlin, J. *Sci. Total Environ.* **2016**, *548–549*, 408–409.

---

6   Mark, H.; Workman, J. *Spectroscopy* **1988**, *3*, 44–48.
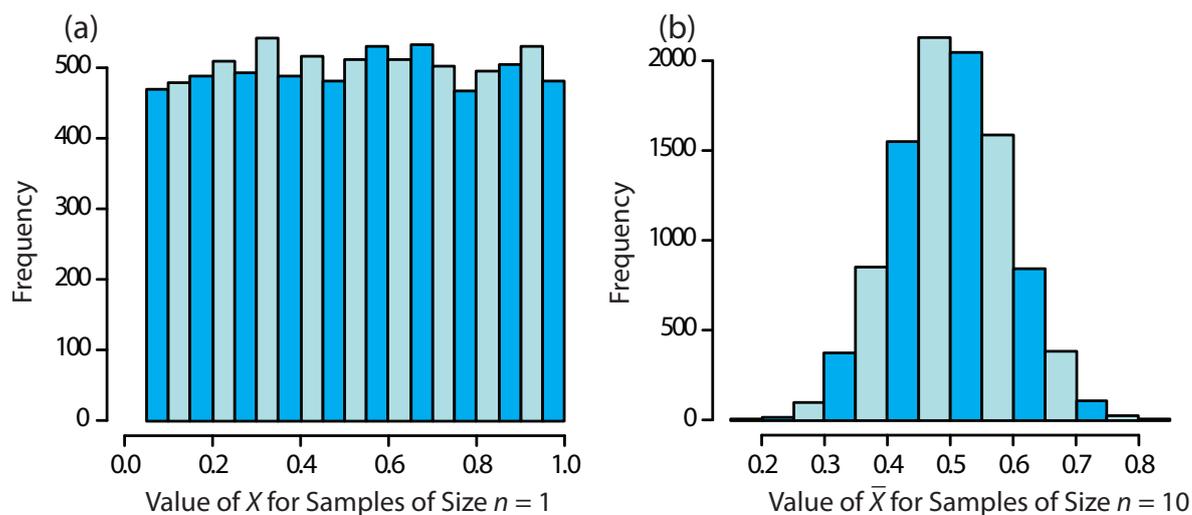
Figure 4.11 Histograms for (a) 10 000 samples of size $n=1$ drawn from a uniform distribution with a minimum value of 0 and a maximum value of 1, and (b) the means for 10 000 samples of size $n=10$ drawn from the same uniform distribution. For (a) the mean of the 10 000 samples is 0.5042, and for (b) the mean of the 10 000 samples is 0.5006. Note that for (a) the distribution closely approximates a uniform distribution in which every possible result is equally likely, and that for (b) the distribution closely approximates a normal distribution.

lation. A more interesting difference is between the denominators of the two equations. When we calculate the population's variance we divide the numerator by the population's size, n; for the sample's variance, however, we divide by $n-1$, where n is the sample's size. Why do we divide by $n-1$ when we calculate the sample's variance?

A variance is the average squared deviation of individual results relative to the mean. When we calculate an average we divide the sum by the number of independent measurements, or DEGREES OF FREEDOM, in the calculation. For the population's variance, the degrees of freedom is equal to the population's size, n. When we measure every member of a population we have complete information about the population.

When we calculate the sample's variance, however, we replace μ with $\overline{X}$, which we also calculate using the same data. If there are n members in the sample, we can deduce the value of the nth member from the remaining $n-1$ members and the mean. For example, if $n=5$ and we know that the first four samples are 1, 2, 3 and 4, and that the mean is 3, then the fifth member of the sample must be

$$X_5 = (\overline{X} \times n) - X_1 - X_2 - X_3 - X_4 =$$
$$(3 \times 5) - 1 - 2 - 3 - 4 = 5$$

Because we have just four independent measurements, we have lost one degree of freedom. Using $n-1$ in place of n when we calculate the sample's variance ensures that s^2 is an unbiased estimator of σ^2.

Here is another way to think about degrees of freedom. We analyze samples to make predictions about the underlying population. When our sample consists of n measurements we cannot make more than n independent predictions about the population. Each time we estimate a parameter, such as the population's mean, we lose a degree of freedom. If there are n degrees of freedom for calculating the sample's mean, then $n-1$ degrees of freedom remain when we calculate the sample's variance.

4D.5 Confidence Intervals for Samples

Earlier we introduced the confidence interval as a way to report the most probable value for a population's mean, μ,

$$\mu = \overline{X} \pm \frac{z\sigma}{\sqrt{n}} \qquad 4.11$$

where $\overline{X}$ is the mean for a sample of size n, and σ is the population's standard deviation. For most analyses we do not know the population's standard deviation. We can still calculate a confidence interval, however, if we make two modifications to equation 4.11.

The first modification is straightforward—we replace the population's standard deviation, σ, with the sample's standard deviation, s. The second modification is not as obvious. The values of z in Table 4.12 are for a normal distribution, which is a function of σ^2, not s^2. Although the sample's variance, s^2, is an unbiased estimate of the population's variance, σ^2, the value of s^2 will only rarely equal σ^2. To account for this uncertainty in estimating σ^2, we replace the variable z in equation 4.11 with the variable t, where t is defined such that $t \geq z$ at all confidence levels.

$$\mu = \overline{X} \pm \frac{ts}{\sqrt{n}} \qquad 4.12$$

Values for t at the 95% confidence level are shown in Table 4.15. Note that t becomes smaller as the number of degrees of freedom increases, and that it approaches z as n approaches infinity. The larger the sample, the more closely its confidence interval for a sample (equation 4.12) approaches the confidence interval for the population (equation 4.11). Appendix 4 provides additional values of t for other confidence levels.

Table 4.15	Values of t for a 95% Confidence Interval		
Degrees of Freedom	t	Degrees of Freedom	t
1	12.706	12	2.179
2	4.303	14	2.145
3	3.181	16	2.120
4	2.776	18	2.101
5	2.571	20	2.086
6	2.447	30	2.042
7	2.365	40	2.021
8	2.306	60	2.000
9	2.262	100	1.984
10	2.228	∞	1.960

Example 4.15

What are the 95% confidence intervals for the two samples of pennies in Table 4.11?

SOLUTION

The mean and the standard deviation for first experiment are, respectively, 3.117 g and 0.051 g. Because the sample consists of seven measurements, there are six degrees of freedom. The value of t from Table 4.15, is 2.447. Substituting into equation 4.12 gives

$$\mu = 3.117 \text{ g} \pm \frac{2.447 \times 0.051 \text{ g}}{\sqrt{7}} = 3.117 \text{ g} \pm 0.047 \text{ g}$$

For the second experiment the mean and the standard deviation are 3.081 g and 0.073 g, respectively, with four degrees of freedom. The 95% confidence interval is

$$\mu = 3.081 \text{ g} \pm \frac{2.776 \times 0.037 \text{ g}}{\sqrt{5}} = 3.081 \text{ g} \pm 0.046 \text{ g}$$

Based on the first experiment, the 95% confidence interval for the population's mean is 3.070–3.164 g. For the second experiment, the 95% confidence interval is 3.035–3.127 g. Although the two confidence intervals are not identical—remember, each confidence interval provides a different estimate for μ—the mean for each experiment is contained within the other experiment's confidence interval. There also is an appreciable overlap of the two confidence intervals. Both of these observations are consistent with samples drawn from the same population.

Note that our comparison of these two confidence intervals at this point is somewhat vague and unsatisfying. We will return to this point in the next section, when we consider a statistical approach to comparing the results of experiments.

Practice Exercise 4.7

What is the 95% confidence interval for the sample of 100 pennies in Table 4.13? The mean and the standard deviation for this sample are 3.095 g and 0.0346 g, respectively. Compare your result to the confidence intervals for the samples of pennies in Table 4.11.

Click here when to review your answer to this exercise.

4D.6   A Cautionary Statement

There is a temptation when we analyze data simply to plug numbers into an equation, carry out the calculation, and report the result. This is never a good idea, and you should develop the habit of reviewing and evaluating your data. For example, if you analyze five samples and report an analyte's mean concentration as 0.67 ppm with a standard deviation of 0.64 ppm, then the 95% confidence interval is

$$\mu = 0.67 \text{ ppm} \pm \frac{2.776 \times 0.64 \text{ ppm}}{\sqrt{5}} = 0.67 \text{ ppm} \pm 0.79 \text{ ppm}$$

This confidence interval estimates that the analyte's true concentration is between –0.12 ppm and 1.46 ppm. Including a negative concentration within the confidence interval should lead you to reevaluate your data or your conclusions. A closer examination of your data may convince you that the standard deviation is larger than expected, making the confidence interval too broad, or you may conclude that the analyte's concentration is too small to report with confidence.

We will return to the topic of detection limits near the end of this chapter.

Here is a second example of why you should closely examine your data: results obtained on samples drawn at random from a normally distributed population must be random. If the results for a sequence of samples show a regular pattern or trend, then the underlying population either is not normally distributed or there is a time-dependent determinate error. For example, if we randomly select 20 pennies and find that the mass of each penny is greater than that for the preceding penny, then we might suspect that our balance is drifting out of calibration.

4E   Statistical Analysis of Data

A confidence interval is a useful way to report the result of an analysis because it sets limits on the expected result. In the absence of determinate error, a confidence interval based on a sample's mean indicates the range of values in which we expect to find the population's mean. When we report a 95% confidence interval for the mass of a penny as $3.117 \text{ g} \pm 0.047 \text{ g}$, for example, we are stating that there is only a 5% probability that the penny's expected mass is less than 3.070 g or more than 3.164 g.

Because a confidence interval is a statement of probability, it allows us to consider comparative questions, such as these: "Are the results for a newly developed method to determine cholesterol in blood significantly different from those obtained using a standard method?" or "Is there a significant variation in the composition of rainwater collected at different sites downwind from a coal-burning utility plant?" In this section we introduce a general approach to the statistical analysis of data. Specific statistical tests are presented in Section 4F.

The reliability of significance testing recently has received much attention—see Nuzzo, R. "Scientific Method: Statistical Errors," *Nature*, **2014**, *506*, 150–152 for a general discussion of the issues—so it is appropriate to begin this section by noting the need to ensure that our data and our research question are compatible so that we do not read more into a statistical analysis than our data allows; see Leek, J. T.; Peng, R. D. "What is the Question? *Science*, **2015**, *347*, 1314-1315 for a useful discussion of six common research questions.

In the context of analytical chemistry, significance testing often accompanies an exploratory data analysis (Is there a reason to suspect that there is a difference between these two analytical methods when applied to a common sample?) or an inferential data analysis (Is there a reason to suspect that there is a relationship between these two independent measurements?). A statistically significant result for these types of analytical research questions generally leads to the design of additional experiments better suited to making predictions or to explaining an underlying causal relationship. A significance test is the first step toward building a greater understanding of an analytical problem, not the final answer to that problem.

4E.1   Significance Testing

Let's consider the following problem. To determine if a medication is effective in lowering blood glucose concentrations, we collect two sets of blood samples from a patient. We collect one set of samples immediately before we administer the medication, and collect the second set of samples several hours later. After analyzing the samples, we report their respective means and variances. How do we decide if the medication was successful in lowering the patient's concentration of blood glucose?

One way to answer this question is to construct a normal distribution curve for each sample, and to compare the two curves to each other. Three

Figure 4.12 Three examples of the possible relationships between the normal distribution curves for two samples. In (a) the curves do not overlap, which suggests that the samples are significantly different from each other. In (b) the two curves are almost identical, suggesting the samples are indistinguishable. The partial overlap of the curves in (c) means that the best we can do is evaluate the probability that there is a difference between the samples.

possible outcomes are shown in Figure 4.12. In Figure 4.12a, there is a complete separation of the two normal distribution curves, which suggests the two samples are significantly different from each other. In Figure 4.12b, the normal distribution curves for the two samples almost completely overlap, which suggests that the difference between the samples is insignificant. Figure 4.12c, however, presents us with a dilemma. Although the means for the two samples seem different, the overlap of their normal distribution curves suggests that a significant number of possible outcomes could belong to either distribution. In this case the best we can do is to make a statement about the probability that the samples are significantly different from each other.

The process by which we determine the probability that there is a significant difference between two samples is called significance testing or hypothesis testing. Before we discuss specific examples we will first establish a general approach to conducting and interpreting a significance test.

4E.2 Constructing a Significance Test

The purpose of a SIGNIFICANCE TEST is to determine whether the difference between two or more results is sufficiently large that it cannot be explained by indeterminate errors. The first step in constructing a significance test is to state the problem as a yes or no question, such as "Is this medication effective at lowering a patient's blood glucose levels?" A null hypothesis and an alternative hypothesis define the two possible answers to our yes or no question. The NULL HYPOTHESIS, H_0, is that indeterminate errors are sufficient to explain any differences between our results. The ALTERNATIVE HYPOTHESIS, H_A, is that the differences in our results are too great to be explained by random error and that they must be determinate in nature. We test the null hypothesis, which we either retain or reject. If we reject the null hypothesis, then we must accept the alternative hypothesis and conclude that the difference is significant.

Failing to reject a null hypothesis is not the same as accepting it. We retain a null hypothesis because we have insufficient evidence to prove it incorrect. It is impossible to prove that a null hypothesis is true. This is an important point and one that is easy to forget. To appreciate this point let's return to our sample of 100 pennies in Table 4.13. After looking at the data we might propose the following null and alternative hypotheses.

H_0: The mass of a circulating U.S. penny is between 2.900 g–3.200 g.

H_A: The mass of a circulating U.S. penny may be less than 2.900 g or more than 3.200 g.

To test the null hypothesis we find a penny and determine its mass. If the penny's mass is 2.512 g then we can reject the null hypothesis and accept the alternative hypothesis. Suppose that the penny's mass is 3.162 g. Although this result increases our confidence in the null hypothesis, it does

not prove that the null hypothesis is correct because the next penny we sample might weigh less than 2.900 g or more than 3.200 g.

After we state the null and the alternative hypotheses, the second step is to choose a confidence level for the analysis. The confidence level defines the probability that we will reject the null hypothesis when it is, in fact, true. We can express this as our confidence that we are correct in rejecting the null hypothesis (e.g. 95%), or as the probability that we are incorrect in rejecting the null hypothesis. For the latter, the confidence level is given as α, where

$$\alpha = 1 - \frac{\text{confidence level (\%)}}{100}$$

For a 95% confidence level, α is 0.05.

The third step is to calculate an appropriate test statistic and to compare it to a critical value. The test statistic's critical value defines a breakpoint between values that lead us to reject or to retain the null hypothesis. How we calculate the test statistic depends on what we are comparing, a topic we cover in section 4F. The last step is to either retain the null hypothesis, or to reject it and accept the alternative hypothesis.

4E.3 One-Tailed and Two-Tailed Significance Tests

Suppose we want to evaluate the accuracy of a new analytical method. We might use the method to analyze a Standard Reference Material that contains a known concentration of analyte, μ. We analyze the standard several times, obtaining a mean value, $\overline{X}$, for the analyte's concentration. Our null hypothesis is that there is no difference between $\overline{X}$ and μ

$$H_0: \overline{X} = \mu$$

If we conduct the significance test at $\alpha = 0.05$, then we retain the null hypothesis if a 95% confidence interval around $\overline{X}$ contains μ. If the alternative hypothesis is

$$H_A: \overline{X} \neq \mu$$

then we reject the null hypothesis and accept the alternative hypothesis if μ lies in the shaded areas at either end of the sample's probability distribution curve (Figure 4.13a). Each of the shaded areas accounts for 2.5% of the area under the probability distribution curve, for a total of 5%. This is a TWO-TAILED SIGNIFICANCE TEST because we reject the null hypothesis for values of μ at either extreme of the sample's probability distribution curve.

We also can write the alternative hypothesis in two additional ways

$$H_A: \overline{X} > \mu$$

$$H_A: \overline{X} < \mu$$

rejecting the null hypothesis if μ falls within the shaded areas shown in Figure 4.13b or Figure 4.13c, respectively. In each case the shaded area

The four steps for a statistical analysis of data using a significance test:

1. Pose a question, and state the null hypothesis, H_0, and the alternative hypothesis, H_A.

3. Choose a confidence level for the statistical analysis.

3. Calculate an appropriate test statistic and compare it to a critical value.

4. Either retain the null hypothesis, or reject it and accept the alternative hypothesis.

In this textbook we use α to represent the probability that we incorrectly reject the null hypothesis. In other textbooks this probability is given as p (often read as "p-value"). Although the symbols differ, the meaning is the same.
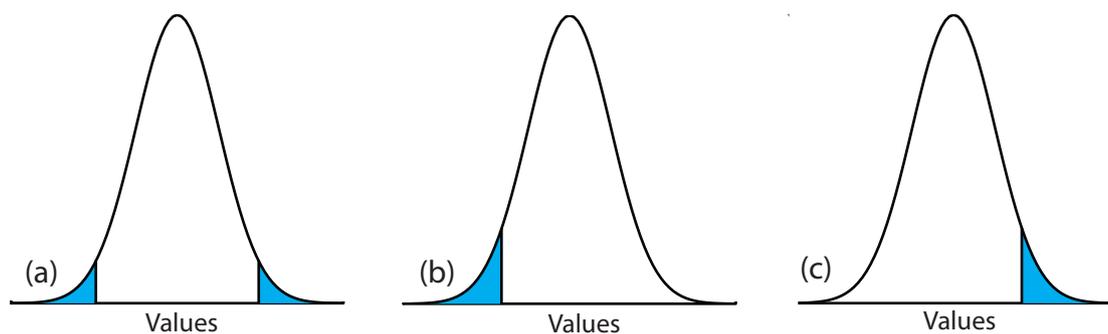
Figure 4.13 Examples of (a) two-tailed, and (b, c) one-tailed, significance test of $\overline{X}$ and μ. The probability distribution curves, which are normal distributions, are based on the sample's mean and standard deviation. For $\alpha = 0.05$, the blue areas account for 5% of the area under the curve. If the value of μ falls within the blue areas, then we reject the null hypothesis and accept the alternative hypothesis. We retain the null hypothesis if the value of μ falls within the unshaded area of the curve.

represents 5% of the area under the probability distribution curve. These are examples of a ONE-TAILED SIGNIFICANCE TEST.

For a fixed confidence level, a two-tailed significance test is the more conservative test because rejecting the null hypothesis requires a larger difference between the parameters we are comparing. In most situations we have no particular reason to expect that one parameter must be larger (or must be smaller) than the other parameter. This is the case, for example, when we evaluate the accuracy of a new analytical method. A two-tailed significance test, therefore, usually is the appropriate choice.

We reserve a one-tailed significance test for a situation where we specifically are interested in whether one parameter is larger (or smaller) than the other parameter. For example, a one-tailed significance test is appropriate if we are evaluating a medication's ability to lower blood glucose levels. In this case we are interested only in whether the glucose levels after we administer the medication is less than the glucose levels before we initiated treatment. If the patient's blood glucose level is greater after we administer the medication, then we know the answer—the medication did not work—and do not need to conduct a statistical analysis.

4E.4   Errors in Significance Testing

Because a significance test relies on probability, its interpretation is subject to error. In a significance test, α defines the probability of rejecting a null hypothesis that is true. When we conduct a significance test at $\alpha = 0.05$, there is a 5% probability that we will incorrectly reject the null hypothesis. This is known as a TYPE 1 ERROR, and its risk is always equivalent to α. A type 1 error in a two-tailed or a one-tailed significance tests corresponds to the shaded areas under the probability distribution curves in Figure 4.13.

A second type of error occurs when we retain a null hypothesis even though it is false. This is as a TYPE 2 ERROR, and the probability of its oc-

currence is β. Unfortunately, in most cases we cannot calculate or estimate the value for β. The probability of a type 2 error, however, is inversely proportional to the probability of a type 1 error.

Minimizing a type 1 error by decreasing α increases the likelihood of a type 2 error. When we choose a value for α we must compromise between these two types of error. Most of the examples in this text use a 95% confidence level ($\alpha = 0.05$) because this usually is a reasonable compromise between type 1 and type 2 errors for analytical work. It is not unusual, however, to use a more stringent (e.g. $\alpha = 0.01$) or a more lenient (e.g. $\alpha = 0.10$) confidence level when the situation calls for it.

4F Statistical Methods for Normal Distributions

The most common distribution for our results is a normal distribution. Because the area between any two limits of a normal distribution curve is well defined, constructing and evaluating significance tests is straightforward.

4F.1 Comparing $\overline{X}$ to μ

One way to validate a new analytical method is to analyze a sample that contains a known amount of analyte, μ. To judge the method's accuracy we analyze several portions of the sample, determine the average amount of analyte in the sample, $\overline{X}$, and use a significance test to compare $\overline{X}$ to μ. Our null hypothesis is that the difference between $\overline{X}$ and μ is explained by indeterminate errors that affect the determination of $\overline{X}$. The alternative hypothesis is that the difference between $\overline{X}$ and μ is too large to be explained by indeterminate error.

$H_0: \overline{X} = \mu$

$H_A: \overline{X} \neq \mu$

The test statistic is t_{exp}, which we substitute into the confidence interval for μ (equation 4.12).

$$\mu = \overline{X} \pm \frac{t_{\text{exp}} s}{\sqrt{n}} \qquad 4.14$$

Rearranging this equation and solving for t_{exp}

$$t_{\text{exp}} = \frac{|\mu - \overline{X}|\sqrt{n}}{s} \qquad 4.15$$

gives the value of t_{exp} when μ is at either the right edge or the left edge of the sample's confidence interval (Figure 4.14a).

To determine if we should retain or reject the null hypothesis, we compare the value of t_{exp} to a critical value, $t(\alpha, \nu)$, where α is the confidence level and ν is the degrees of freedom for the sample. The critical value $t(\alpha, \nu)$ defines the largest confidence interval explained by indeterminate error. If $t_{\text{exp}} > t(\alpha, \nu)$, then our sample's confidence interval is greater than that explained by indeterminate errors (Figure 4.14b). In this case, we reject the null hypothesis and accept the alternative hypothesis. If $t_{\text{exp}} \leq t(\alpha, \nu)$, then our sample's confidence interval is smaller than that explained by inde-

Values for $t(\alpha, \nu)$ are in Appendix 4.
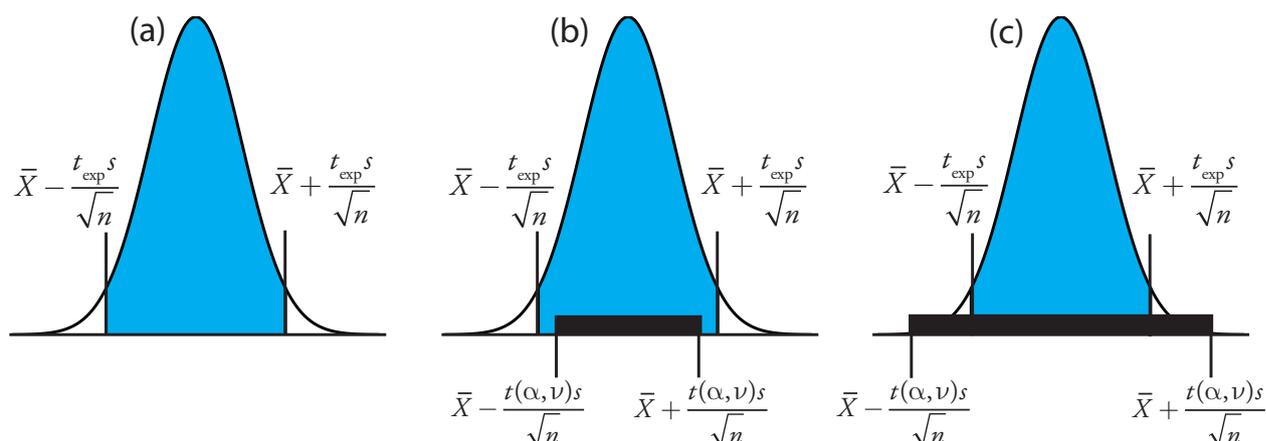
Figure 4.14 Relationship between a confidence interval and the result of a significance test. (a) The shaded area under the normal distribution curve shows the sample's confidence interval for μ based on t_{exp}. The solid bars in (b) and (c) show the expected confidence intervals for μ explained by indeterminate error given the choice of α and the available degrees of freedom, ν. For (b) we reject the null hypothesis because portions of the sample's confidence interval fall outside the confidence interval explained by indeterminate error. In the case of (c) we retain the null hypothesis because the confidence interval explained by indeterminate error completely encompasses the sample's confidence interval.

Another name for the t-test is Student's t-test. Student was the pen name for William Gossett (1876-1927) who developed the t-test while working as a statistician for the Guiness Brewery in Dublin, Ireland. He published under the name Student because the brewery did not want its competitors to know they were using statistics to help improve the quality of their products.

terminate error, and we retain the null hypothesis (Figure 4.14c). Example 4.16 provides a typical application of this significance test, which is known as a t-TEST of $\overline{X}$ to μ.

Example 4.16

Before determining the amount of Na_2CO_3 in a sample, you decide to check your procedure by analyzing a standard sample that is 98.76% w/w Na_2CO_3. Five replicate determinations of the %w/w Na_2CO_3 in the standard gave the following results.

$$98.71\% \quad 98.59\% \quad 98.62\% \quad 98.44\% \quad 98.58\%$$

Using $\alpha = 0.05$, is there any evidence that the analysis is giving inaccurate results?

Solution

The mean and standard deviation for the five trials are

$$\overline{X} = 98.59 \qquad\qquad s = 0.0973$$

Because there is no reason to believe that the results for the standard must be larger or smaller than μ, a two-tailed t-test is appropriate. The null hypothesis and alternative hypothesis are

$$H_0: \overline{X} = \mu \qquad\qquad H_A: \overline{X} \neq \mu$$

The test statistic, t_{exp}, is

$$t_{exp} = \frac{|\mu - \overline{X}|\sqrt{n}}{s} = \frac{|98.76 - 98.59|\sqrt{5}}{0.0973} = 3.91$$

The critical value for $t(0.05,4)$ from Appendix 4 is 2.78. Since t_{exp} is greater than $t(0.05, 4)$, we reject the null hypothesis and accept the alternative hypothesis. At the 95% confidence level the difference between $\overline{X}$ and μ is too large to be explained by indeterminate sources of error, which suggests there is a determinate source of error that affects the analysis.

Practice Exercise 4.8

To evaluate the accuracy of a new analytical method, an analyst determines the purity of a standard for which μ is 100.0%, obtaining the following results.

 99.28% 103.93% 99.43% 99.84% 97.60% 96.70% 98.02%

Is there any evidence at $\alpha = 0.05$ that there is a determinate error affecting the results?

Click here to review your answer to this exercise.

Earlier we made the point that we must exercise caution when we interpret the result of a statistical analysis. We will keep returning to this point because it is an important one. Having determined that a result is inaccurate, as we did in Example 4.16, the next step is to identify and to correct the error. Before we expend time and money on this, however, we first should examine critically our data. For example, the smaller the value of s, the larger the value of t_{exp}. If the standard deviation for our analysis is unrealistically small, then the probability of a type 2 error increases. Including a few additional replicate analyses of the standard and reevaluating the t-test may strengthen our evidence for a determinate error, or it may show us that there is no evidence for a determinate error.

4F.2   Comparing s^2 to σ^2

If we analyze regularly a particular sample, we may be able to establish an expected variance, σ^2, for the analysis. This often is the case, for example, in a clinical lab that analyze hundreds of blood samples each day. A few replicate analyses of a single sample gives a sample variance, s^2, whose value may or may not differ significantly from σ^2.

We can use an *F*-TEST to evaluate whether a difference between s^2 and σ^2 is significant. The null hypothesis is $H_0: s^2 = \sigma^2$ and the alternative hypothesis is $H_A: s^2 \neq \sigma^2$. The test statistic for evaluating the null hypothesis is F_{exp}, which is given as either

$$F_{exp} = \frac{s^2}{\sigma^2} \text{ if } s^2 > \sigma^2 \text{ or } F_{exp} = \frac{\sigma^2}{s^2} \text{ if } \sigma^2 > s^2 \qquad 4.16$$

depending on whether s^2 is larger or smaller than σ^2. This way of defining F_{exp} ensures that its value is always greater than or equal to one.

If the null hypothesis is true, then F_{exp} should equal one; however, because of indeterminate errors F_{exp} usually is greater than one. A critical value, $F(\alpha, \nu_{num}, \nu_{den})$, is the largest value of F_{exp} that we can attribute to indeterminate error given the specified significance level, α, and the degrees of freedom for the variance in the numerator, ν_{num}, and the variance in the denominator, ν_{den}. The degrees of freedom for s^2 is $n-1$, where n is the number of replicates used to determine the sample's variance, and the degrees of freedom for σ^2 is defined as infinity, ∞. Critical values of F for $\alpha = 0.05$ are listed in Appendix 5 for both one-tailed and two-tailed F-tests.

Example 4.17

A manufacturer's process for analyzing aspirin tablets has a known variance of 25. A sample of 10 aspirin tablets is selected and analyzed for the amount of aspirin, yielding the following results in mg aspirin/tablet.

<div align="center">254   249   252   252   249   249   250   247   251   252</div>

Determine whether there is evidence of a significant difference between the sample's variance and the expected variance at $\alpha=0.05$.

SOLUTION

The variance for the sample of 10 tablets is 4.3. The null hypothesis and alternative hypotheses are

$$H_0: s^2 = \sigma^2 \qquad H_A: s^2 \neq \sigma^2$$

and the value for F_{exp} is

$$F_{exp} = \frac{\sigma^2}{s^2} = \frac{25}{4.3} = 5.8$$

The critical value for $F(0.05, \infty, 9)$ from Appendix 5 is 3.333. Since F_{exp} is greater than $F(0.05, \infty, 9)$, we reject the null hypothesis and accept the alternative hypothesis that there is a significant difference between the sample's variance and the expected variance. One explanation for the difference might be that the aspirin tablets were not selected randomly.

4F.3 Comparing Two Sample Variances

We can extend the F-test to compare the variances for two samples, A and B, by rewriting equation 4.16 as

$$F_{exp} = \frac{s_A^2}{s_B^2}$$

defining A and B so that the value of F_{exp} is greater than or equal to 1.

Example 4.18

Table 4.11 shows results for two experiments to determine the mass of a circulating U.S. penny. Determine whether there is a difference in the variances of these analyses at $\alpha = 0.05$.

SOLUTION

The variances for the two experiments are 0.00259 for the first experiment (A) and 0.00138 for the second experiment (B). The null and alternative hypotheses are

$$H_0: s_A^2 = s_B^2 \qquad H_A: s_A^2 \neq s_B^2$$

and the value of F_{exp} is

$$F_{exp} = \frac{s_A^2}{s_B^2} = \frac{(0.051)^2}{(0.037)^2} = \frac{0.00260}{0.00137} = 1.90$$

From Appendix 5, the critical value for $F(0.05, 6, 4)$ is 9.197. Because $F_{exp} < F(0.05, 6, 4)$, we retain the null hypothesis. There is no evidence at $\alpha = 0.05$ to suggest that the difference in variances is significant.

Practice Exercise 4.9

To compare two production lots of aspirin tablets, we collect ana analyze samples from each, obtaining the following results (in mg aspirin/tablet).

Lot 1: 256   248   245   245   244   248   261

Lot 2: 241   258   241   244   256   254

Is there any evidence at $\alpha = 0.05$ that there is a significant difference in the variances for these two samples?

Click here to review your answer to this exercise.

4F.4   Comparing Two Sample Means

Three factors influence the result of an analysis: the method, the sample, and the analyst. We can study the influence of these factors by conducting experiments in which we change one factor while holding constant the other factors. For example, to compare two analytical methods we can have the same analyst apply each method to the same sample and then examine the resulting means. In a similar fashion, we can design experiments to compare two analysts or to compare two samples.

Before we consider the significance tests for comparing the means of two samples, we need to make a distinction between unpaired data and paired data. This is a critical distinction and learning to distinguish between these two types of data is important. Here are two simple examples that highlight the difference between UNPAIRED DATA and PAIRED DATA. In each example the goal is to compare two balances by weighing pennies.

It also is possible to design experiments in which we vary more than one of these factors. We will return to this point in Chapter 14.

- Example 1: We collect 10 pennies and weigh each penny on each balance. This is an example of paired data because we use the same 10 pennies to evaluate each balance.

- Example 2: We collect 10 pennies and divide them into two groups of five pennies each. We weight the pennies in the first group on one balance and we weigh the second group of pennies on the other balance. Note that no penny is weighed on both balances. This is an example of unpaired data because we evaluate each balance using a different sample of pennies.

In both examples the samples of 10 pennies were drawn from the same population; the difference is how we sampled that population. We will learn why this distinction is important when we review the significance test for paired data; first, however, we present the significance test for unpaired data.

UNPAIRED DATA

Consider two analyses, A and B with means of $\overline{X}_A$ and $\overline{X}_B$, and standard deviations of s_A and s_B. The confidence intervals for μ_A and for μ_B are

$$\mu_A = \overline{X}_A \pm \frac{ts_A}{\sqrt{n_A}}$$
4.17

$$\mu_B = \overline{X}_B \pm \frac{ts_B}{\sqrt{n_B}}$$
4.18

where n_A and n_B are the sample sizes for A and for B. Our null hypothesis, $H_0: \mu_A = \mu_B$, is that and any difference between μ_A and μ_B is the result of indeterminate errors that affect the analyses. The alternative hypothesis, $H_A: \mu_A \neq \mu_B$, is that the difference between μ_A and μ_B is too large to be explained by indeterminate error.

To derive an equation for t_{exp}, we assume that μ_A equals μ_B, and combine equations 4.17 and 4.18.

$$\overline{X}_A \pm \frac{t_{exp}s_A}{\sqrt{n_A}} = \overline{X}_B \pm \frac{t_{exp}s_B}{\sqrt{n_B}}$$

Solving for $|\overline{X}_A - \overline{X}_B|$ and using a propagation of uncertainty, gives

$$|\overline{X}_A - \overline{X}_B| = t_{exp} \times \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$
4.19

Finally, we solve for t_{exp}

$$t_{exp} = \frac{|\overline{X}_A - \overline{X}_B|}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$
4.20

and compare it to a critical value, $t(\alpha, \nu)$, where α is the probability of a type 1 error, and ν is the degrees of freedom.

Thus far our development of this t-test is similar to that for comparing $\overline{X}$ to μ, and yet we do not have enough information to evaluate the t-test.

Do you see the problem? With two independent sets of data it is unclear how many degrees of freedom we have.

Suppose that the variances s_A^2 and s_B^2 provide estimates of the same σ^2. In this case we can replace s_A^2 and s_B^2 with a pooled variance, s_{pool}^2, that is a better estimate for the variance. Thus, equation 4.20 becomes

$$t_{exp} = \frac{|\overline{X}_A - \overline{X}_B|}{s_{pool} \times \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{|\overline{X}_A - \overline{X}_B|}{s_{pool}} \times \sqrt{\frac{n_A n_B}{n_A + n_B}} \qquad 4.21$$

where s_{pool}, the pooled standard deviation, is

$$s_{pool} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}} \qquad 4.22$$

So how do you determine if it is okay to pool the variances? Use an *F*-test.

The denominator of equation 4.22 shows us that the degrees of freedom for a pooled standard deviation is $n_A + n_B - 2$, which also is the degrees of freedom for the *t*-test. Note that we lose two degrees of freedom because the calculations for s_A^2 and s_B^2 require the prior calculation of $\overline{X}_A$ and $\overline{X}_B$.

If s_A^2 and s_B^2 are significantly different, then we calculate t_{exp} using equation 4.20. In this case, we find the degrees of freedom using the following imposing equation.

$$\nu = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{\left(\frac{s_A^2}{n_A}\right)^2}{n_A + 1} + \frac{\left(\frac{s_B^2}{n_B}\right)^2}{n_B + 1}} - 2 \qquad 4.23$$

Because the degrees of freedom must be an integer, we round to the nearest integer the value of ν obtained using equation 4.23.

Regardless of whether we calculate t_{exp} using equation 4.20 or equation 4.21, we reject the null hypothesis if t_{exp} is greater than $t(\alpha, \nu)$ and retain the null hypothesis if t_{exp} is less than or equal to $t(\alpha, \nu)$.

Example 4.19

Tables 4.11 provides results for two experiments to determine the mass of a circulating U.S. penny. Determine whether there is a difference in the means of these analyses at $\alpha = 0.05$.

SOLUTION

First we use an *F*-test to determine whether we can pool the variances. We completed this analysis in Example 4.18, finding no evidence of a significant difference, which means we can pool the standard deviations, obtaining

$$s_{pool} = \sqrt{\frac{(7 - 1)(0.051)^2 + (5 - 1)(0.037)^2}{7 + 5 - 2}} = 0.0459$$

with 10 degrees of freedom. To compare the means we use the following null hypothesis and alternative hypotheses

$$H_0: \mu_A = \mu_B \qquad H_A: \mu_A \neq \mu_B$$

Because we are using the pooled standard deviation, we calculate t_{exp} using equation 4.21.

$$t_{exp} = \frac{|3.117 - 3.081|}{0.0459} \times \sqrt{\frac{7 \times 5}{7 + 5}} = 1.34$$

The critical value for $t(0.05, 10)$, from Appendix 4, is 2.23. Because t_{exp} is less than $t(0.05, 10)$ we retain the null hypothesis. For $\alpha = 0.05$ we do not have evidence that the two sets of pennies are significantly different.

Example 4.20

One method for determining the %w/w Na_2CO_3 in soda ash is to use an acid–base titration. When two analysts analyze the same sample of soda ash they obtain the results shown here.

Analyst A	Analyst B
86.82	81.01
87.04	86.15
86.93	81.73
87.01	83.19
86.20	80.27
87.00	83.93

Determine whether the difference in the mean values is significant at $\alpha = 0.05$.

SOLUTION

We begin by reporting the mean and standard deviation for each analyst.

$$\overline{X}_A = 86.83\% \qquad s_A = 0.32$$
$$\overline{X}_B = 82.71\% \qquad s_A = 2.16$$

To determine whether we can use a pooled standard deviation, we first complete an F-test using the following null and alternative hypotheses.

$$H_0: s_A^2 = s_B^2 \qquad H_A: s_A^2 \neq s_B^2$$

Calculating F_{exp}, we obtain a value of

$$F_{exp} = \frac{(2.16)^2}{(0.32)^2} = 45.6$$

Because F_{exp} is larger than the critical value of 7.15 for $F(0.05, 5, 5)$ from Appendix 5, we reject the null hypothesis and accept the alternative hypothesis that there is a significant difference between the variances; thus, we cannot calculate a pooled standard deviation.

To compare the means for the two analysts we use the following null and alternative hypotheses.

$$H_0: \mu_A = \mu_B \qquad H_A: \mu_A \neq \mu_B$$

Because we cannot pool the standard deviations, we calculate t_{exp} using equation 4.20 instead of equation 4.21

$$t_{exp} = \frac{|86.83 - 82.71|}{\sqrt{\dfrac{(0.32)^2}{6} + \dfrac{(2.16)^2}{6}}} = 4.62$$

and calculate the degrees of freedom using equation 4.23.

$$\nu = \frac{\left(\dfrac{(0.32)^2}{6} + \dfrac{(2.16)^2}{6}\right)^2}{\dfrac{\left(\dfrac{(0.32)^2}{6}\right)^2}{6+1} + \dfrac{\left(\dfrac{(2.16)^2}{6}\right)^2}{6+1}} - 2 = 5.3 \approx 5$$

From Appendix 4, the critical value for $t(0.05, 5)$ is 2.57. Because t_{exp} is greater than $t(0.05, 5)$ we reject the null hypothesis and accept the alternative hypothesis that the means for the two analysts are significantly different at $\alpha = 0.05$.

Practice Exercise 4.10

To compare two production lots of aspirin tablets, you collect samples from each and analyze them, obtaining the following results (in mg aspirin/tablet).

Lot 1: 256   248   245   245   244   248   261

Lot 2: 241   258   241   244   256   254

Is there any evidence at $\alpha = 0.05$ that there is a significant difference in the variance between the results for these two samples? This is the same data from Practice Exercise 4.9.

Click here to review your answer to this exercise.

PAIRED DATA

Suppose we are evaluating a new method for monitoring blood glucose concentrations in patients. An important part of evaluating a new method is to compare it to an established method. What is the best way to gather data for this study? Because the variation in the blood glucose levels amongst patients is large we may be unable to detect a small, but significant difference between the methods if we use different patients to gather data for each method. Using paired data, in which the we analyze each patient's blood using both methods, prevents a large variance within a population from adversely affecting a t-test of means.

Typical blood glucose levels for most non-diabetic individuals ranges between 80–120 mg/dL (4.4–6.7 mM), rising to as high as 140 mg/dL (7.8 mM) shortly after eating. Higher levels are common for individuals who are pre-diabetic or diabetic.

When we use paired data we first calculate the difference, d_i, between the paired values for each sample. Using these difference values, we then calculate the average difference, $\overline{d}$, and the standard deviation of the differences, s_d. The null hypothesis, $H_0: \overline{d} = 0$, is that there is no difference between the two samples, and the alternative hypothesis, $H_A: \overline{d} \neq 0$, is that the difference between the two samples is significant.

The test statistic, t_{exp}, is derived from a confidence interval around $\overline{d}$

$$ t_{exp} = \frac{|\overline{d}|\sqrt{n}}{s_d} $$

where n is the number of paired samples. As is true for other forms of the t-test, we compare t_{exp} to $t(\alpha, \nu)$, where the degrees of freedom, ν, is $n-1$. If t_{exp} is greater than $t(\alpha, \nu)$, then we reject the null hypothesis and accept the alternative hypothesis. We retain the null hypothesis if t_{exp} is less than or equal to $t(\alpha, \nu)$. This is known as a PAIRED t-TEST.

Example 4.21

Marecek et. al. developed a new electrochemical method for the rapid determination of the concentration of the antibiotic monensin in fermentation vats.[7] The standard method for the analysis is a test for microbiological activity, which is both difficult to complete and time-consuming. Samples were collected from the fermentation vats at various times during production and analyzed for the concentration of monensin using both methods. The results, in parts per thousand (ppt), are reported in the following table.

Sample	Microbiological	Electrochemical
1	129.5	132.3
2	89.6	91.0
3	76.6	73.6
4	52.2	58.2
5	110.8	104.2
6	50.4	49.9
7	72.4	82.1
8	141.4	154.1
9	75.0	73.4
10	34.1	38.1
11	60.3	60.1

Is there a significant difference between the methods at $\alpha = 0.05$?

SOLUTION

Acquiring samples over an extended period of time introduces a substantial time-dependent change in the concentration of monensin. Because the

7   Marecek, V.; Janchenova, H.; Brezina, M.; Betti, M. *Anal. Chim. Acta* **1991**, *244*, 15–19.

variation in concentration between samples is so large, we use a paired t-test with the following null and alternative hypotheses.

$$H_0: \overline{d} = 0 \qquad H_A: \overline{d} \neq 0$$

Defining the difference between the methods as

$$d_i = (X_{elect})_i - (X_{micro})_i$$

we calculate the difference for each sample.

Sample	1	2	3	4	5	6	7	8	9	10	11
d_i	2.8	1.4	–3.0	6.0	–6.6	–0.5	9.7	12.7	–1.6	4.0	–0.2

The mean and the standard deviation for the differences are, respectively, 2.25 ppt and 5.63 ppt. The value of t_{exp} is

$$t_{exp} = \frac{|2.25|\sqrt{11}}{5.63} = 1.33$$

which is smaller than the critical value of 2.23 for $t(0.05, 10)$ from [Appendix 4](#). We retain the null hypothesis and find no evidence for a significant difference in the methods at $\alpha = 0.05$.

One important requirement for a paired t-test is that the determinate and the indeterminate errors that affect the analysis must be independent of the analyte's concentration. If this is not the case, then a sample with an unusually high concentration of analyte will have an unusually large d_i. Including this sample in the calculation of $\overline{d}$ and s_d gives a biased estimate for the expected mean and standard deviation. This rarely is a problem for samples that span a limited range of analyte concentrations, such as those in [Example 4.21](#) or Practice Exercise 4.11. When paired data span a wide range of concentrations, however, the magnitude of the determinate and indeterminate sources of error may not be independent of the analyte's con-

---

Practice Exercise 4.11

Suppose you are studying the distribution of zinc in a lake and want to know if there is a significant difference between the concentration of Zn^{2+} at the sediment-water interface and its concentration at the air-water interface. You collect samples from six locations—near the lake's center, near its drainage outlet, etc.—obtaining the results (in mg/L) shown in the table. Using this data, determine if there is a significant difference between the concentration of Zn^{2+} at the two interfaces at $\alpha = 0.05$.

Location	Air-Water Interface	Sediment-Water Interface
1	0.430	0.415
2	0.266	0.238
3	0.457	0.390
4	0.531	0.410
5	0.707	0.605
6	0.716	0.609

Complete this analysis treating the data as (a) unpaired and as (b) paired. Briefly comment on your results.

Click [here](#) to review your answers to this exercise.

centration; when true, a paired *t*-test may give misleading results because the paired data with the largest absolute determinate and indeterminate errors will dominate $\bar{d}$. In this situation a regression analysis, which is the subject of the next chapter, is more appropriate method for comparing the data.

4F.5   Outliers

Earlier in the chapter we examined several data sets consisting of the mass of a circulating United States penny. Table 4.16 provides one more data set. Do you notice anything unusual in this data? Of the 112 pennies included in Table 4.11 and Table 4.13, no penny weighed less than 3 g. In Table 4.16, however, the mass of one penny is less than 3 g. We might ask whether this penny's mass is so different from the other pennies that it is in error.

A measurement that is not consistent with other measurements is called outlier. An OUTLIER might exist for many reasons: the outlier might belong to a different population (Is this a Canadian penny?); the outlier might be a contaminated or otherwise altered sample (Is the penny damaged or unusually dirty?); or the outlier may result from an error in the analysis (Did we forget to tare the balance?). Regardless of its source, the presence of an outlier compromises any meaningful analysis of our data. There are many significance tests that we can use to identify a potential outlier, three of which we present here.

DIXON'S *Q*-TEST

One of the most common significance tests for identifying an outlier is DIXON'S *Q*-TEST. The null hypothesis is that there are no outliers, and the alternative hypothesis is that there is an outlier. The *Q*-test compares the gap between the suspected outlier and its nearest numerical neighbor to the range of the entire data set (Figure 4.15). The test statistic, Q_{exp}, is

$$Q_{exp} = \frac{\text{gap}}{\text{range}} = \frac{|\text{outlier's value} - \text{nearest value}|}{\text{largest value} - \text{smallest value}}$$

This equation is appropriate for evaluating a single outlier. Other forms of Dixon's *Q*-test allow its extension to detecting multiple outliers.[8]

The value of Q_{exp} is compared to a critical value, $Q(\alpha, n)$, where α is the probability that we will reject a valid data point (a type 1 error) and n is the total number of data points. To protect against rejecting a valid data point, usually we apply the more conservative two-tailed *Q*-test, even though the

---

8   Rorabacher, D. B. *Anal. Chem.* **1991**, *63*, 139–146.

Table 4.16 Mass (g) for Additional Sample of Circulating U. S. Pennies		
3.067	2.514	3.094
3.049	3.048	3.109
3.039	3.079	3.102

(a)

(b)
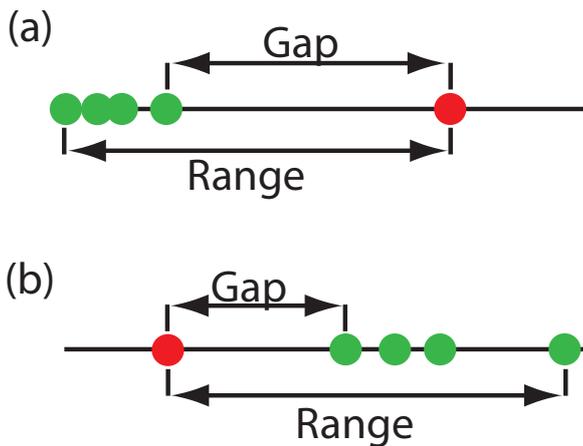
Figure 4.15 Dotplots showing the distribution of two data sets containing a possible outlier. In (a) the possible **outlier's** value is larger than the remaining **data**, and in (b) the possible **outlier's** value is smaller than the remaining **data**.

possible outlier is the smallest or the largest value in the data set. If Q_{exp} is greater than $Q(\alpha, n)$, then we reject the null hypothesis and may exclude the outlier. We retain the possible outlier when Q_{exp} is less than or equal to $Q(\alpha, n)$. Table 4.17 provides values for $Q(0.05, n)$ for a data set that has 3–10 values. A more extensive table is in Appendix 6. Values for $Q(\alpha, n)$ assume an underlying normal distribution.

GRUBB'S TEST

Although Dixon's Q-test is a common method for evaluating outliers, it is no longer favored by the International Standards Organization (ISO), which recommends GRUBB'S TEST.[9] There are several versions of Grubb's test depending on the number of potential outliers. Here we will consider the case where there is a single suspected outlier.

The test statistic for Grubb's test, G_{exp}, is the distance between the sample's mean, $\overline{X}$, and the potential outlier, X_{out}, in terms of the sample's standard deviation, s.

$$G_{exp} = \frac{|X_{out} - \overline{X}|}{s}$$

We compare the value of G_{exp} to a critical value $G(\alpha, n)$, where α is the probability that we will reject a valid data point and n is the number of data points in the sample. If G_{exp} is greater than $G(\alpha, n)$, then we may reject the data point as an outlier, otherwise we retain the data point as part of the sample. Table 4.18 provides values for $G(0.05, n)$ for a sample containing 3–10 values. A more extensive table is in Appendix 7. Values for $G(\alpha, n)$ assume an underlying normal distribution.

Table 4.17 Dixon's Q-Test	
n	$Q(0.05, n)$
3	0.970
4	0.829
5	0.710
6	0.625
7	0.568
8	0.526
9	0.493
10	0.466

Table 4.18 Grubb's Test	
n	$G(0.05, n)$
3	1.115
4	1.481
5	1.715
6	1.887
7	2.020
8	2.126
9	2.215
10	2.290

9   International Standards ISO Guide 5752-2 "Accuracy (trueness and precision) of measurement methods and results–Part 2: basic methods for the determination of repeatability and reproducibility of a standard measurement method," 1994.

CHAUVENET'S CRITERION

Our final method for identifying an outlier is CHAUVENET'S CRITERION. Unlike Dixon's Q-Test and Grubb's test, you can apply this method to any distribution as long as you know how to calculate the probability for a particular outcome. Chauvenet's criterion states that we can reject a data point if the probability of obtaining the data point's value is less than $(2n)^{-1}$, where n is the size of the sample. For example, if $n = 10$, a result with a probability of less than $(2 \times 10)^{-1}$, or 0.05, is considered an outlier.

To calculate a potential outlier's probability we first calculate its standardized deviation, z

$$z = \frac{|X_{out} - \overline{X}|}{s}$$

where X_{out} is the potential outlier, $\overline{X}$ is the sample's mean and s is the sample's standard deviation. Note that this equation is identical to the equation for G_{exp} in the Grubb's test. For a normal distribution, we can find the probability of obtaining a value of z using the probability table in Appendix 3.

Example 4.22

Table 4.16 contains the masses for nine circulating United States pennies. One entry, 2.514 g, appears to be an outlier. Determine if this penny is an outlier using a Q-test, Grubb's test, and Chauvenet's criterion. For the Q-test and Grubb's test, let $\alpha = 0.05$.

SOLUTION

For the Q-test the value for Q_{exp} is

$$Q_{exp} = \frac{|2.514 - 3.039|}{3.109 - 2.514} = 0.882$$

From Table 4.17, the critical value for $Q(0.05, 9)$ is 0.493. Because Q_{exp} is greater than $Q(0.05, 9)$, we can assume the penny with a mass of 2.514 g likely is an outlier.

For Grubb's test we first need the mean and the standard deviation, which are 3.011 g and 0.188 g, respectively. The value for G_{exp} is

$$G_{exp} = \frac{|2.514 - 3.011|}{0.188} = 2.64$$

Using Table 4.18, we find that the critical value for $G(0.05, 9)$ is 2.215. Because G_{exp} is greater than $G(0.05, 9)$, we can assume that the penny with a mass of 2.514 g likely is an outlier.

For Chauvenet's criterion, the critical probability is $(2 \times 9)^{-1}$, or 0.0556. The value of z is the same as G_{exp}, or 2.64. Using Appendix 3, the probability for $z = 2.64$ is 0.00415. Because the probability of obtaining a mass of 0.2514 g is less than the critical probability, we can assume the penny with a mass of 2.514 g likely is an outlier.

You should exercise caution when using a significance test for outliers because there is a chance you will reject a valid result. In addition, you should avoid rejecting an outlier if it leads to a precision that is much better than expected based on a propagation of uncertainty. Given these concerns it is not surprising that some statisticians caution against the removal of outliers.[10]

On the other hand, testing for outliers can provide useful information if we try to understand the source of the suspected outlier. For example, the outlier in Table 4.16 represents a significant change in the mass of a penny (an approximately 17% decrease in mass), which is the result of a change in the composition of the U.S. penny. In 1982 the composition of a U.S. penny changed from a brass alloy that was 95% w/w Cu and 5% w/w Zn (with a nominal mass of 3.1 g), to a pure zinc core covered with copper (with a nominal mass of 2.5 g).[11] The pennies in Table 4.16, therefore, were drawn from different populations.

4G   Detection Limits

The International Union of Pure and Applied Chemistry (IUPAC) defines a method's DETECTION LIMIT as the smallest concentration or absolute amount of analyte that has a signal significantly larger than the signal from a suitable blank.[12] Although our interest is in the amount of analyte, in this section we will define the detection limit in terms of the analyte's signal. Knowing the signal you can calculate the analyte's concentration, C_A, or the moles of analyte, n_A, using the equations

$$S_A = k_A C_A \quad \text{or} \quad S_A = k_A n_A$$

where k is the method's sensitivity.

Let's translate the IUPAC definition of the detection limit into a mathematical form by letting S_{mb} represent the average signal for a method blank, and letting σ_{mb} represent the method blank's standard deviation. The null hypothesis is that the analyte is not present in the sample, and the alternative hypothesis is that the analyte is present in the sample. To detect the analyte, its signal must exceed S_{mb} by a suitable amount; thus,

$$(S_A)_{\text{DL}} = S_{mb} + z\sigma_{mb} \qquad 4.24$$

where $(S_A)_{\text{DL}}$ is the analyte's detection limit.

The value we choose for z depends on our tolerance for reporting the analyte's concentration even if it is absent from the sample (a type 1 error). Typically, z is set to three, which, from Appendix 3, corresponds to a probability, α, of 0.00135. As shown in Figure 4.16a, there is only a 0.135% probability of detecting the analyte in a sample that actually is analyte-free.

You also can adopt a more stringent requirement for rejecting data. When using the Grubb's test, for example, the ISO 5752 guidelines suggests retaining a value if the probability for rejecting it is greater than $\alpha = 0.05$, and flagging a value as a "straggler" if the probability for rejecting it is between $\alpha = 0.05$ and 0.01. A "straggler" is retained unless there is compelling reason for its rejection. The guidelines recommend using $\alpha = 0.01$ as the minimum criterion for rejecting a possible outlier.

See Chapter 3 for a review of these equations.

If σ_{mb} is not known, we can replace it with s_{mb}; equation 4.24 then becomes

$$(S_A)_{\text{DL}} = S_{mb} \pm t s_{mb}$$

You can make similar adjustments to other equations in this section.

See, for example, Kirchner, C. J. "Estimation of Detection Limits for Environmental Analytical Procedures," in Currie, L. A. (ed) *Detection in Analytical Chemistry: Importance, Theory, and Practice*; American Chemical Society: Washington, D. C., 1988.

---

10 Deming, W. E. *Statistical Analysis of Data*; Wiley: New York, 1943 (republished by Dover: New York, 1961); p. 171.
11 Richardson, T. H. *J. Chem. Educ.* **1991**, *68*, 310–311.
12 IUPAC Compendium of Chemical Technology, Electronic Version, http://goldbook.iupac.org/D01629.html
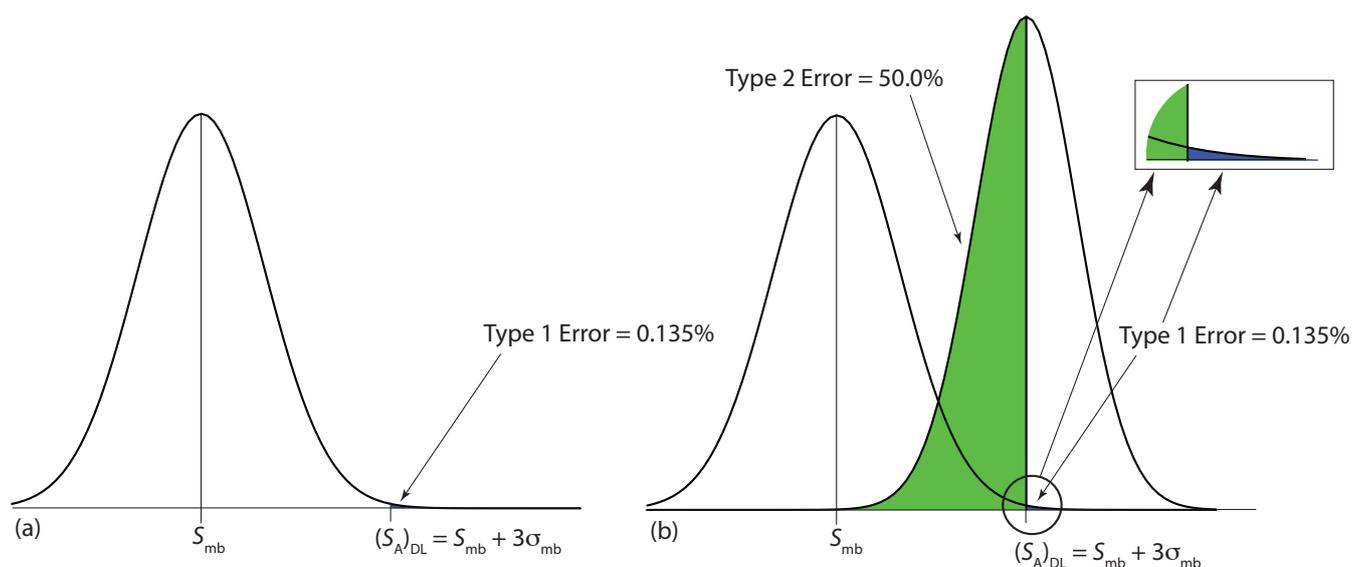
Figure 4.16 Normal distribution curves showing the probability of type 1 and type 2 errors for the IUPAC detection limit. (a) The normal distribution curve for the method blank, with $S_{mb}=0$ and $\sigma_{mb}=1$. The minimum detectable signal for the analyte, $(S_A)_{DL}$, has a type 1 error of 0.135%. (b) The normal distribution curve for the analyte at its detection limit, $(S_A)_{DL}=3$, is superimposed on the normal distribution curve for the method blank. The standard deviation for the analyte's signal, σ_A, is 0.8, The area in **green** represents the probability of a type 2 error, which is 50%. The inset shows, in **blue**, the probability of a type 1 error, which is 0.135%.

A detection limit also is subject to a type 2 error in which we fail to find evidence for the analyte even though it is present in the sample. Consider, for example, the situation shown in Figure 4.16b where the signal for a sample that contains the analyte is exactly equal to $(S_A)_{DL}$. In this case the probability of a type 2 error is 50% because half of the sample's possible signals are below the detection limit. We correctly detect the analyte at the IUPAC detection limit only half the time. The IUPAC definition for the detection limit is the smallest signal for which we can say, at a significance level of α, that an analyte is present in the sample; however, failing to detect the analyte does not mean it is not present in the sample.

The detection limit often is represented, particularly when discussing public policy issues, as a distinct line that separates detectable concentrations of analytes from concentrations we cannot detect. This use of a detection limit is incorrect.[13] As suggested by Figure 4.16, for an analyte whose concentration is near the detection limit there is a high probability that we will fail to detect the analyte.

An alternative expression for the detection limit, the LIMIT OF IDENTIFICATION, minimizes both type 1 and type 2 errors.[14] The analyte's signal at the limit of identification, $(S_A)_{LOI}$, includes an additional term, $z\sigma_A$, to account for the distribution of the analyte's signal.

$$(S_A)_{LOI} = (S_A)_{DL} + z\sigma_A = S_{mb} + z\sigma_{mb} + z\sigma_A$$

13 Rogers, L. B. *J. Chem. Educ.* **1986**, *63*, 3–6.
14 Long, G. L.; Winefordner, J. D. *Anal. Chem.* **1983**, *55*, 712A–724A.
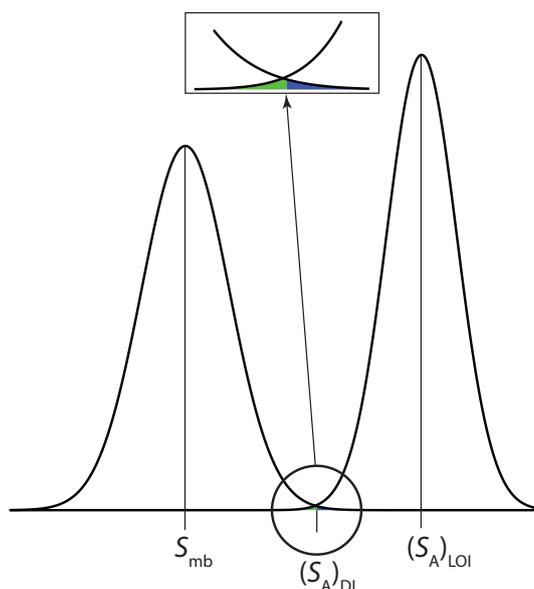
Figure 4.17 Normal distribution curves for a method blank and for a sample at the limit of identification: $S_{mb} = 0$; $\sigma_{mb} = 1$; $\sigma_A = 0.8$; and $(S_A)_{\text{LOI}} = 0 + 3 \times 1 + 3 \times 0.8 = 5.4$. The inset shows that the probability of a **type 1 error** (0.135%) is the same as the probability of a **type 2 error** (0.135%).

As shown in Figure 4.17, the limit of identification provides an equal probability of a type 1 and a type 2 error at the detection limit. When the analyte's concentration is at its limit of identification, there is only a 0.135% probability that its signal is indistinguishable from that of the method blank.

The ability to detect the analyte with confidence is not the same as the ability to report with confidence its concentration, or to distinguish between its concentration in two samples. For this reason the American Chemical Society's Committee on Environmental Analytical Chemistry recommends the LIMIT OF QUANTITATION, $(S_A)_{\text{LOQ}}$.[15]

$$(S_A)_{\text{LOQ}} = S_{mb} + 10\sigma_{mb}$$

4H   Using Excel and R to Analyze Data

Although the calculations in this chapter are relatively straightforward, it can be tedious to work problems using nothing more than a calculator. Both Excel and R include functions for many common statistical calculations. In addition, R provides useful functions for visualizing your data.

4H.1   Excel

Excel has built-in functions that we can use to complete many of the statistical calculations covered in this chapter, including reporting descriptive statistics, such as means and variances, predicting the probability of obtaining a given outcome from a binomial distribution or a normal distribution, and carrying out significance tests. Table 4.19 provides the syntax for many of these functions; you can information on functions not included here by using Excel's Help menu.

---

15   "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," *Anal. Chem.* **1980**, *52*, 2242–2249.

Table 4.19 Excel Functions for Statistics Calculations	
Parameter	Excel Function
Descriptive Statistics	
mean	=average(data)
median	=median(data)
sample standard deviation	=stdev.s(data)
population standard deviation	=stdev.p(data)
sample variance	=var.s(data)
population variance	=var.p(data)
maximum value	=max(data)
minimum value	=min(data)
Probability Distributions	
binomial distribution	=binom.dist(X, N, p, TRUE or FALSE)
normal distribution	=norm.dist(x, μ, σ, TRUE or FALSE)
Significance Tests	
F-test	f.test(data set 1, data set 2)
t-test	t.test(data set 1, data set 2, tails =1 or 2, type of t-test: 1 = paired; 2 = unpaired with equal variances; or 3 = unpaired with unequal variances)

DESCRIPTIVE STATISTICS

Let's use Excel to provide a statistical summary of the data in Table 4.1. Enter the data into a spreadsheet, as shown in Figure 4.18. To calculate the sample's mean, for example, click on any empty cell, enter the formula

$$=\textbf{average}(b2:b8)$$

and press Return or Enter to replace the cell's content with Excel's calculation of the mean (3.117 285 714), which we round to 3.117. Excel does not have a function for the range, but we can use the functions that report the maximum value and the minimum value to calculate the range; thus

$$=\textbf{max}(b2:b8) - \textbf{min}(b2:b8)$$

returns 0.142 as an answer.

PROBABILITY DISTRIBUTIONS

In Example 4.11 we showed that 91.10% of a manufacturer's analgesic tablets contained between 243 and 262 mg of aspirin. We arrived at this result by calculating the deviation, z, of each limit from the population's

	A	B
1		mass (g)
2		3.080
3		3.094
4		3.107
5		3.056
6		3.112
7		3.174
8		3.198

Figure 4.18 Portion of a spreadsheet containing data from Table 4.1.

expected mean, μ, of 250 mg in terms of the population's expected standard deviation, σ, of 5 mg. After we calculated values for z, we used the table in Appendix 3 to find the area under the normal distribution curve between these two limits.

We can complete this calculation in Excel using the **norm.dist** function As shown in Figure 4.19, the function calculates the probability of obtaining a result less than x from a normal distribution with a mean of μ and a standard deviation of σ. To solve Example 4.11 using Excel enter the following formulas into separate cells

$$=\textbf{norm.dist}(243, 250, 5, \text{TRUE})$$

$$=\textbf{norm.dist}(262, 250, 5, \text{TRUE})$$

obtaining results of 0.080 756 659 and 0.991 802 464. Subtracting the smaller value from the larger value and adjusting to the correct number of significant figures gives the probability as 0.9910, or 99.10%.

Excel also includes a function for working with binomial distributions. The function's syntax is

$$=\textbf{binom.dist}(X, N, p, \text{TRUE or FALSE})$$

where X is the number of times a particular outcome occurs in N trials, and p is the probability that X occurs in a single trial. Setting the function's last term to TRUE gives the total probability for any result up to X and setting it to FALSE gives the probability for X. Using Example 4.10 to test this function, we use the formula

$$=\textbf{binom.dist}(0, 27, 0.0111, \text{FALSE})$$

to find the probability of finding no atoms of ^{13}C atoms in a molecule of cholesterol, $C_{27}H_{44}O$, which returns a value of 0.740 after adjusting for significant figures. Using the formula

$$=\textbf{binom.dist}(2, 27, 0.0111, \text{TRUE})$$

we find that 99.7% of cholesterol molecules contain two or fewer atoms of ^{13}C.

SIGNIFICANCE TESTS

As shown in Table 4.19, Excel includes functions for the following significance tests covered in this chapter:

- an F-test of variances
- an unpaired t-test of sample means assuming equal variances
- an unpaired t-test of sample means assuming unequal variances
- a paired t-test for of sample means

Let's use these functions to complete a t-test on the data in Table 4.11, which contains results for two experiments to determine the mass of a circulating U. S. penny. Enter the data from Table 4.11 into a spreadsheet
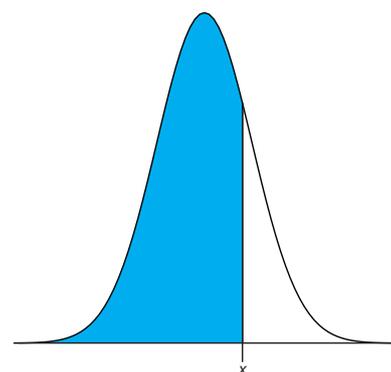
Figure 4.19 Shown in blue is the area returned by the function

$$=\textbf{norm.dist}(x, \mu, \sigma, \text{TRUE})$$

The last parameter—TRUE—returns the cumulative distribution from $-\infty$ to x; entering FALSE gives the probability of obtaining the result x. For our purposes, we want to use TRUE.

	A	B	C
1		Set 1	Set 2
2		3.080	3.052
3		3.094	3.141
4		3.107	3.083
5		3.056	3.083
6		3.112	3.048
7		3.174	
8		3.198	

Figure 4.20 Portion of a spreadsheet containing the data in Table 4.11.

See Example 4.18 and Example 4.19 for our earlier solutions to this problem.

as shown in Figure 4.20. Because the data in this case are unpaired, we will use Excel to complete an unpaired t-test. Before we can complete the t-test, we use an F-test to determine whether the variances for the two data sets are equal or unequal.

To complete the F-test, we click on any empty cell, enter the formula

$$=\textbf{f.test}(b2:b8, c2:c6)$$

and press Return or Enter, which replaces the cell's content with the value of α for which we can reject the null hypothesis of equal variances. In this case, Excel returns an α of 0.566 105 03; because this value is not less than 0.05, we retain the null hypothesis that the variances are equal. Excel's F-test is two-tailed; for a one-tailed F-test, we use the same function, but divide the result by two; thus

$$=\textbf{f.test}(b2:b8, c2:c6)/2$$

Having found no evidence to suggest unequal variances, we next complete an unpaired t-test assuming equal variances, entering into any empty cell the formula

$$=\textbf{t.test}(b2:b8, c2:c6, 2, 2)$$

where the first 2 indicates that this is a two-tailed t-test, and the second 2 indicates that this is an unpaired t-test with equal variances. Pressing Return or Enter replaces the cell's content with the value of α for which we can reject the null hypothesis of equal means. In this case, Excel returns an α of 0.211 627 646; because this value is not less than 0.05, we retain the null hypothesis that the means are equal.

The other significance tests in Excel work in the same format. The following practice exercise provides you with an opportunity to test yourself.

Practice Exercise 4.12

Rework Example 4.20 and Example 4.21 using Excel.

Click here to review your answers to this exercise.

4H.2   R

You can download the current version of R from www.r-project.org. Click on the link for Download: CRAN and find a local mirror site. Click on the link for the mirror site and then use the link for Linux, MacOS X, or Windows under the heading "Download and Install R."

R is a programming environment that provides powerful capabilities for analyzing data. There are many functions built into R's standard installation and additional packages of functions are available from the R web site (www.r-project.org). Commands in R are not available from pull down menus. Instead, you interact with R by typing in commands.

DESCRIPTIVE STATISTICS

Let's use R to provide a statistical summary of the data in Table 4.1. To do this we first need to create an object that contains the data, which we do by typing in the following command.

Table 4.20 R Functions for Descriptive Statistics	
Parameter	**R Function**
mean	mean(object)
median	median(object)
sample standard deviation	sd(object)
population standard deviation	sd(object) * ((length(object)–1)/length(object))^0.5
sample variance	var(object)
population variance	var(object) * ((length(object)–1)/length(object))
range	max(object) – min(object)

> penny1 = c(3.080, 3.094, 3.107, 3.056, 3.112, 3.174, 3.198)

Table 4.20 lists some of the commands in R for calculating basic descriptive statistics. As is the case for Excel, R does not include stand alone commands for all descriptive statistics of interest to us, but we can calculate them using other commands. Using a command is easy—simply enter the appropriate code at the prompt; for example, to find the sample's variance we enter

> **var**(penny1)

[1] 0.002221918

PROBABILITY DISTRIBUTIONS

In Example 4.11 we showed that 91.10% of a manufacturer's analgesic tablets contain between 243 and 262 mg of aspirin. We obtained this result by calculating the deviation, z, of each limit from the population's expected mean, μ, of 250 mg in terms of the population's expected standard deviation, σ, of 5 mg. After we calculated values for z, we used the table in Appendix 3 to find the area under the normal distribution curve between the two limits.

We can complete this calculation in R using the function **pnorm**. The function's general format is

$$\textbf{pnorm}(x, \mu, \sigma)$$

where x is the limit of interest, μ is the distribution's expected mean, and σ is the distribution's expected standard deviation. The function returns the probability of obtaining a result of less than x (Figure 4.21). Here is the output of an R session for solving Example 4.11.

> pnorm(243, 250, 5)

[1] 0.08075666

> pnorm(262, 250, 5)

[1] 0.9918025

In R, the symbol '>' is a prompt, which indicates that the program is waiting for you to enter a command. When you press 'Return' or 'Enter,' R executes the command, displays the result (if there is a result to return), and returns the > prompt.
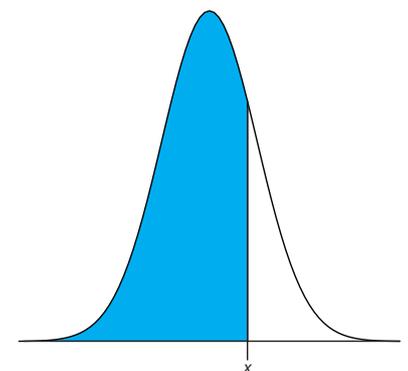
Figure 4.21 Shown in blue is the area returned by the function

$$\text{pnorm}(x, \mu, \sigma)$$

Subtracting the smaller value from the larger value and adjusting to the correct number of significant figures gives the probability as 0.9910, or 99.10%.

R also includes functions for binomial distributions. To find the probability of obtaining a particular outcome, X, in N trials we use the **dbinom** function.

$$\textbf{dbinom}(X, N, p)$$

where X is the number of times a particular outcome occurs in N trials, and p is the probability that X occurs in a single trial. Using Example 4.10 to test this function, we find that the probability of finding no atoms of ^{13}C atoms in a molecule of cholesterol, $C_{27}H_{44}O$ is

> dbinom(0, 27, 0.0111)

[1] 0.7397997

0.740 after adjusting the significant figures. To find the probability of obtaining any outcome up to a maximum value of X, we use the **pbinom** function.

$$\textbf{pbinom}(X, N, p)$$

To find the percentage of cholesterol molecules that contain 0, 1, or 2 atoms of ^{13}C, we enter

> pbinom(2, 27, 0.0111)

[1] 0.9967226

and find that the answer is 99.7% of cholesterol molecules.

Significance Tests

R includes commands for the following significance tests covered in this chapter:

- F-test of variances
- unpaired t-test of sample means assuming equal variances
- unpaired t-test of sample means assuming unequal variances
- paired t-test for of sample means
- Dixon's Q-test for outliers
- Grubb's test for outliers

Let's use R to complete a t-test on the data in Table 4.11, which contains results for two experiments to determine the mass of a circulating U. S. penny. First, enter the data from Table 4.11 into two objects.

> penny1 = c(3.080, 3.094, 3.107, 3.056, 3.112, 3.174, 3.198)

> penny2 = c(3.052, 3.141, 3.083, 3.083, 3.048)

Because the data in this case are unpaired, we will use R to complete an unpaired t-test. Before we can complete a t-test we use an F-test to determine whether the variances for the two data sets are equal or unequal.

To complete a two-tailed F-test in R we use the command

<div align="center">

var.test(X, Y)

</div>

where X and Y are the objects that contain the two data sets. Figure 4.22 shows the output from an R session to solve this problem. Note that R does not provide the critical value for $F(0.05, 6, 4)$; instead it reports the 95% confidence interval for F_{exp}. Because this confidence interval of 0.204 to 11.661 includes the expected value for F of 1.00, we retain the null hypothesis and have no evidence for a difference between the variances. R also provides the probability of incorrectly rejecting the null hypothesis, which in this case is 0.5561.

Having found no evidence suggesting unequal variances, we now complete an unpaired t-test assuming equal variances. The basic syntax for a two-tailed t-test is

<div align="center">

t.test(X, Y, <u>mu = 0, paired = FALSE, var.equal = FALSE</u>)

</div>

where X and Y are the objects that contain the data sets. You can change the underlined terms to alter the nature of the t-test. Replacing "var.equal = FALSE" to "var.equal = TRUE" makes this a two-tailed t-test with equal variances, and replacing "paired = FALSE" with "paired = TRUE" makes this a paired t-test. The term "mu = 0" is the expected difference between the means, which for this problem is 0. You can, of course, change this to suit your needs. The underlined terms are default values; if you omit them, then R assumes you intend an unpaired two-tailed t-test of the null hypothesis that $X = Y$ with unequal variances. Figure 4.23 shows the output of an R session for this problem.

We can interpret the results of this t-test in two ways. First, the p-value of 0.2116 means there is a 21.16% probability of incorrectly rejecting the

> var.test(penny1, penny2)

   F test to compare two variances

data: penny1 and penny2

F = 1.8726, num df = 6, denom df = 4, p-value = 0.5661
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.2036028 11.6609726
sample estimates:
ratio of variances
      1.872598

Figure 4.22 Output of an R session for an F-test of variances. The p-value of 0.5661 is the probability of incorrectly rejecting the null hypothesis that the variances are equal (*note: R identifies* α *as a p-value*). The 95% confidence interval is the range of values for F_{exp} that are explained by random error. If this range includes the expected value for F, in this case 1.00, then there is insufficient evidence to reject the null hypothesis. Note that R does not adjust for significant figures.

> t.test(penny1, penny2, var.equal=TRUE)

  Two Sample t-test

data: penny1 and penny2

t = 1.3345, df = 10, p-value = 0.2116
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02403040 0.09580182
sample estimates:
mean of x mean of y
 3.117286 3.081400

Figure 4.23 Output of an R session for an unpaired *t*-test with equal variances. The *p*-value of 0.2116 is the probability of incorrectly rejecting the null hypothesis that the means are equal (*note: R identifies α as a p-value*). The 95% confidence interval is the range of values for the difference between the means that is explained by random error. If this range includes the expected value for the difference, in this case zero, then there is insufficient evidence to reject the null hypothesis. Note that R does not adjust for significant figures.

null hypothesis. Second, the 95% confidence interval of -0.024 to 0.0958 for the difference between the sample means includes the expected value of zero. Both ways of looking at the results provide no evidence for rejecting the null hypothesis; thus, we retain the null hypothesis and find no evidence for a difference between the two samples.

The other significance tests in R work in the same format. The following practice exercise provides you with an opportunity to test yourself.

Practice Exercise 4.13

Rework Example 4.20 and Example 4.21 using R.

Click here to review your answers to this exercise.

Unlike Excel, R also includes functions for evaluating outliers. These functions are not part of R's standard installation. To install them enter the following command within R (*note: you will need an internet connection to download the package of functions*).

> install.packages("outliers")

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

> library("outliers")

Let's use this package to find the outlier in Table 4.16 using both Dixon's *Q*-test and Grubb's test. The commands for these tests are

dixon.test(*X*, type = 10, two.sided = TRUE)

grubbs.test(X, type = 10, two.sided = TRUE)

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).

```
> penny3=c(3.067,3.049, 3.039, 2.514, 3.048, 3.079, 3.094, 3.109, 3.102)
> dixon.test(penny3, type=10, two.sided=TRUE)

    Dixon test for outliers

data:  penny3
Q = 0.8824, p-value < 2.2e-16
alternative hypothesis: lowest value 2.514 is an outlier

> grubbs.test(penny3, type=10, two.sided=TRUE)

    Grubbs test for one outlier

data:  penny3
G = 2.6430, U = 0.0177, p-value = 1.938e-06
alternative hypothesis: lowest value 2.514 is an outlier
```

Figure 4.24 Output of an R session for Dixon's *Q*-test and Grubb's test for outliers. The *p*-values for both tests show that we can treat as an outlier the penny with a mass of 2.514 g.

where X is the object that contains the data, "type $= 10$" specifies that we are looking for one outlier, and "two.sided$=$TRUE" indicates that we are using the more conservative two-tailed test. Both tests have other variants that allow for the testing of outliers on both ends of the data set ("type$= 11$") or for more than one outlier ("type$= 20$"), but we will not consider these here. Figure 4.24 shows the output of a session for this problem. For both tests the very small *p*-value indicates that we can treat as an outlier the penny with a mass of 2.514 g.

VISUALIZING DATA

One of R's more useful features is the ability to visualize data. Visualizing data is important because it provides us with an intuitive feel for our data that can help us in applying and evaluating statistical tests. It is tempting to believe that a statistical analysis is foolproof, particularly if the probability for incorrectly rejecting the null hypothesis is small. Looking at a visual display of our data, however, can help us determine whether our data is normally distributed—a requirement for most of the significance tests in this chapter—and can help us identify potential outliers. There are many useful ways to look at data, four of which we consider here.

Visualizing data is important, a point we will return to in Chapter 5 when we consider the mathematical modeling of data.

To plot data in R, we will use the package "lattice," which you will need to load using the following command.

```
> library("lattice")
```

You do not need to use the command install.package this time because lattice was automatically installed on your computer when you downloaded R.

To demonstrate the types of plots we can generate, we will use the object "penny," which contains the masses of the 100 pennies in Table 4.13.

Our first visualization is a histogram. To construct the histogram we use mass to divide the pennies into bins and plot the number of pennies or the percent of pennies in each bin on the *y*-axis as a function of mass on the *x*-axis. Figure 4.25a shows the result of entering the command

To create a histogram showing the number of pennies in each bin, change "percent" to "count."

> **histogram**(penny, type = "percent", xlab = "Mass (g)",
ylab = "Percent of Pennies", main = "Histogram of Data in Table 4.13")

A histogram allows us to visualize the data's distribution. In this example the data appear to follow a normal distribution, although the largest bin does not include the mean of 3.095 g and the distribution is not perfectly symmetric. One limitation of a histogram is that its appearance

(a) Histogram of Data in Table 4.13

(b) Kernel Density Plot of Data in Table 4.13

(c) Dotchart of Data in Table 4.13

(d) Boxplot of Data in Table 4.13
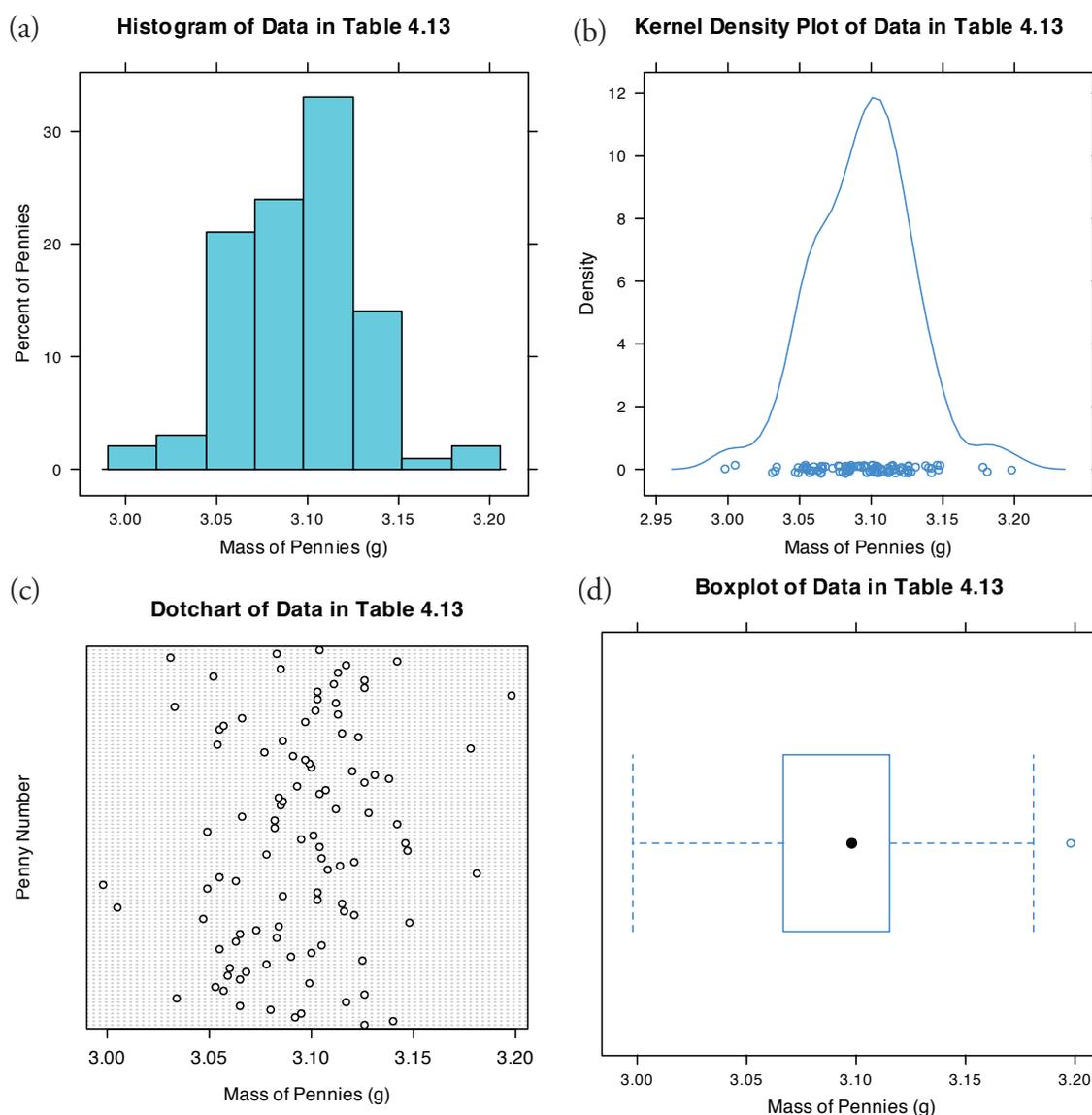
Figure 4.25 Four ways to plot the data in Table 4.13: (a) histogram; (b) kernel density plot showing smoothed distribution and individual data points; (c) dot chart; and (d) box plot.

depends on how we choose to bin the data. Increasing the number of bins and centering the bins around the data's mean gives a histogram that more closely approximates a normal distribution (Figure 4.10).

An alternative to the histogram is a KERNEL DENSITY PLOT, which basically is a smoothed histogram. In this plot each value in the data set is replaced with a normal distribution curve whose width is a function of the data set's standard deviation and size. The resulting curve is a summation of the individual distributions. Figure 4.25b shows the result of entering the command

> **densityplot**(penny, xlab = "Mass of Pennies (g)", main = "Kernel Density Plot of Data in Table 4.13")

The circles at the bottom of the plot show the mass of each penny in the data set. This display provides a more convincing picture that the data in Table 4.13 are normally distributed, although we see evidence of a small clustering of pennies with a mass of approximately 3.06 g.

We analyze samples to characterize the parent population. To reach a meaningful conclusion about a population, the samples must be representative of the population. One important requirement is that the samples are random. A DOT CHART provides a simple visual display that allows us to examine the data for non-random trends. Figure 4.25c shows the result of entering

> **dotchart**(penny, xlab = "Mass of Pennies (g)", ylab = "Penny Number", main = "Dotchart of Data in Table 4.13")

In this plot the masses of the 100 pennies are arranged along the *y*-axis in the order in which they were sampled. If we see a pattern in the data along the *y*-axis, such as a trend toward smaller masses as we move from the first penny to the last penny, then we have clear evidence of non-random sampling. Because our data do not show a pattern, we have more confidence in the quality of our data.

The last plot we will consider is a BOX PLOT, which is a useful way to identify potential outliers without making any assumptions about the data's distribution. A box plot contains four pieces of information about a data set: the median, the middle 50% of the data, the smallest value and the largest value within a set distance of the middle 50% of the data, and possible outliers. Figure 4.25d shows the result of entering

> **bwplot**(penny, xlab = "Mass of Pennies (g)", main = "Boxplot of Data in Table 4.13")"

The black dot (●) is the data set's median. The rectangular box shows the range of masses spanning the middle 50% of the pennies. This also is known as the interquartile range, or IQR. The dashed lines, which are called "whiskers," extend to the smallest value and the largest value that are within ±1.5×IQR of the rectangular box. Potential outliers are shown as open circles (o). For normally distributed data the median is near the

Note that the dispersion of points along the *x*-axis is not uniform, with more points occurring near the center of the *x*-axis than at either end. This pattern is as expected for a normal distribution.

To find the interquartile range you first find the median, which divides the data in half. The median of each half provides the limits for the box. The IQR is the median of the upper half of the data minus the median for the lower half of the data. For the data in Table 4.13 the median is 3.098. The median for the lower half of the data is 3.068 and the median for the upper half of the data is 3.115. The IQR is $3.115 - 3.068 = 0.047$. You can use the command "summary(penny)" in R to obtain these values.

The lower "whisker" extend to the first data point with a mass larger than

$$3.068 - 1.5 \times IQR = 3.068 - 1.5 \times 0.047$$
$$= 2.9975$$

which for this data is 2.998 g. The upper "whisker" extends to the last data point with a mass smaller than

$$3.115 + 1.5 \times IQR = 3.115 + 1.5 \times 0.047 = 3.1855$$

which for this data is 3.181 g.

center of the box and the whiskers will be equidistant from the box. As is often the case in statistics, the converse is not true—finding that a boxplot is perfectly symmetric does not prove that the data are normally distributed.

The box plot in Figure 4.25d is consistent with the histogram (Figure 4.25a) and the kernel density plot (Figure 4.28b). Together, the three plots provide evidence that the data in Table 4.13 are normally distributed. The potential outlier, whose mass of 3.198 g, is not sufficiently far away from the upper whisker to be of concern, particularly as the size of the data set ($n = 100$) is so large. A Grubb's test on the potential outlier does not provide evidence for treating it as an outlier.

> ### Practice Exercise 4.14
>
> Use R to create a data set consisting of 100 values from a uniform distribution by entering the command
>
> > data = runif(100, min = 0, max = 100)
>
> A uniform distribution is one in which every value between the minimum and the maximum is equally probable. Examine the data set by creating a histogram, a kernel density plot, a dot chart, and a box plot. Briefly comment on what the plots tell you about the your sample and its parent population.
>
> Click here to review your answer to this exercise.

4I   Key Terms

alternative hypothesis	bias	binomial distribution
box plot	central limit theorem	Chauvenet's criterion
confidence interval	constant determinate error	degrees of freedom
detection limit	determinate error	Dixon's Q-test
dot chart	error	F-test
Grubb's test	histogram	indeterminate error
kernel density plot	limit of identification	limit of quantitation
mean	median	measurement error
method error	normal distribution	null hypothesis
one-tailed significance test	outlier	paired data
paired t-test	personal error	population
probability distribution	propagation of uncertainty	proportional determinate error
range	repeatability	reproducibility
sample	sampling error	significance test
standard deviation	standard error of the mean	Standard Reference Material
tolerance	t-test	two-tailed significance test
type 1 error	type 2 error	uncertainty
unpaired data	variance	

4J   Chapter Summary

The data we collect are characterized by their central tendency (where the values cluster), and their spread (the variation of individual values around the central value). We report our data's central tendency by stating the mean or median, and our data's spread using the range, standard deviation or variance. Our collection of data is subject to errors, including determinate errors that affect the data's accuracy and indeterminate errors that affect its precision. A propagation of uncertainty allows us to estimate how these determinate and indeterminate errors affect our results.

When we analyze a sample several times the distribution of the results is described by a probability distribution, two examples of which are the binomial distribution and the normal distribution. Knowing the type of distribution allows us to determine the probability of obtaining a particular range of results. For a normal distribution we express this range as a confidence interval.

A statistical analysis allows us to determine whether our results are significantly different from known values, or from values obtained by other analysts, by other methods of analysis, or for other samples. We can use a *t*-test to compare mean values and an *F*-test to compare variances. To compare two sets of data you first must determine whether the data is paired or unpaired. For unpaired data you also must decide if you can pool the standard deviations. A decision about whether to retain an outlying value can be made using Dixon's Q-test, Grubb's test, or Chauvenet's criterion. You should be sure to exercise caution if you decide to reject an outlier.

Finally, the detection limit is a statistical statement about the smallest amount of analyte we can detect with confidence. A detection limit is not exact since its value depends on how willing we are to falsely report the analyte's presence or absence in a sample. When reporting a detection limit you should clearly indicate how you arrived at its value.

4K   Problems

1. The following masses were recorded for 12 different U.S. quarters (all given in grams):

5.683	5.549	5.548	5.552
5.620	5.536	5.539	5.684
5.551	5.552	5.554	5.632

    Report the mean, median, range, standard deviation and variance for this data.

2. A determination of acetaminophen in 10 separate tablets of Excedrin Extra Strength Pain Reliever gives the following results (in mg).[16]

Many of the problems that follow require access to statistical tables. For your convenience, here are hyperlinks to the appendices containing these tables.

Appendix 3: Single-Sided Normal Distribution
Appendix 4: Critical Values for the *t*-Test
Appendix 5: Critical Values for the *F*-Test
Appendix 6: Critical Values for Dixon's *Q*-Test
Appendix 7: Critical Values for Grubb's Test

16 Simonian, M. H.; Dinh, S.; Fray, L. A. *Spectroscopy* **1993**, *8(6)*, 37–47.

| 224.3 | 240.4 | 246.3 | 239.4 | 253.1 |
| 261.7 | 229.4 | 255.5 | 235.5 | 249.7 |

(a) Report the mean, median, range, standard deviation and variance for this data. (b) Assuming that $\overline{X}$ and s^2 are good approximations for μ and for σ^2, and that the population is normally distributed, what percentage of tablets contain more than the standard amount of 250 mg acetaminophen per tablet?

3. Salem and Galan developed a new method to determine the amount of morphine hydrochloride in tablets.[17] An analysis of tablets with different nominal dosages gave the following results (in mg/tablet).

100-mg tablets	60-mg tablets	30-mg tablets	10-mg tablets
99.17	54.21	28.51	9.06
94.31	55.62	26.25	8.83
95.92	57.40	25.92	9.08
94.55	57.51	28.62	
93.83	52.59	24.93	

(a) For each dosage, calculate the mean and the standard deviation for the mg of morphine hydrochloride per tablet. (b) For each dosage level, and assuming that $\overline{X}$ and s^2 are good approximations for μ and for σ^2, and that the population is normally distributed, what percentage of tablets contain more than the nominal amount of morphine hydrochloride per tablet?

4. Daskalakis and co-workers evaluated several procedures for digesting oyster and mussel tissue prior to analyzing them for silver.[18] To evaluate the procedures they spiked samples with known amounts of silver and analyzed the samples to determine the amount of silver, reporting results as the percentage of added silver found in the analysis. A procedure was judged acceptable if its spike recoveries fell within the range $100 \pm 15\%$. The spike recoveries for one method are shown here.

See Chapter 15 to learn more about using a spike recovery to evaluate an analytical method.

| 106% | 108% | 92% | 99% |
| 101% | 93% | 93% | 104% |

Assuming a normal distribution for the spike recoveries, what is the probability that any single spike recovery is within the accepted range?

5. The formula weight (*FW*) of a gas can be determined using the following form of the ideal gas law

17 Salem, I. I.; Galan, A. C. Anal. Chim. Acta 1993, 283, 334–337.

18 Daskalakis, K. D.; O'Connor, T. P.; Crecelius, E. A. *Environ. Sci. Technol.* **1997**, *31*, 2303–2306.

$$FW = \frac{g\mathrm{R}T}{PV}$$

where g is the mass in grams, R is the gas constant, T is the temperature in Kelvin, P is the pressure in atmospheres, and V is the volume in liters. In a typical analysis the following data are obtained (with estimated uncertainties in parentheses)

$g = 0.118$ g $(\pm 0.002$ g$)$

$R = 0.082056$ L atm mol^{-1} K^{-1} $(\pm 0.000001$ L atm mol^{-1} K$^{-1})$

$T = 298.2$ K $(\pm 0.1$ K$)$

$P = 0.724$ atm $(\pm 0.005$ atm$)$

$V = 0.250$ L $(\pm 0.005$ L$)$

(a) What is the compound's formula weight and its estimated uncertainty? (b) To which variable(s) should you direct your attention if you wish to improve the uncertainty in the compound's molecular weight?

6. To prepare a standard solution of Mn^{2+}, a 0.250 g sample of Mn is dissolved in 10 mL of concentrated HNO_3 (measured with a graduated cylinder). The resulting solution is quantitatively transferred to a 100-mL volumetric flask and diluted to volume with distilled water. A 10-mL aliquot of the solution is pipeted into a 500-mL volumetric flask and diluted to volume. (a) Express the concentration of Mn in mg/L, and estimate its uncertainty using a propagation of uncertainty. (b) Can you improve the concentration's uncertainty by using a pipet to measure the HNO_3, instead of a graduated cylinder?

7. The mass of a hygroscopic compound is measured using the technique of weighing by difference. In this technique the compound is placed in a sealed container and weighed. A portion of the compound is removed and the container and the remaining material are reweighed. The difference between the two masses gives the sample's mass. A solution of a hygroscopic compound with a gram formula weight of 121.34 g/mol (± 0.01 g/mol) is prepared in the following manner. A sample of the compound and its container has a mass of 23.5811 g. A portion of the compound is transferred to a 100-mL volumetric flask and diluted to volume. The mass of the compound and container after the transfer is 22.1559 g. Calculate the compound's molarity and estimate its uncertainty by a propagation of uncertainty.

8. Use a propagation of uncertainty to show that the standard error of the mean for n determinations is $\sigma/\sqrt{n}$.

9. Beginning with equation 4.17 and equation 4.18, use a propagation of uncertainty to derive equation 4.19.

10. What is the smallest mass you can measure on an analytical balance that has a tolerance of ± 0.1 mg, if the relative error must be less than 0.1%?

11. Which of the following is the best way to dispense 100.0 mL if we wish to minimize the uncertainty: (a) use a 50-mL pipet twice; (b) use a 25-mL pipet four times; or (c) use a 10-mL pipet ten times?

12. You can dilute a solution by a factor of 200 using readily available pipets (1-mL to 100-mL) and volumetric flasks (10-mL to 1000-mL) in either one step, two steps, or three steps. Limiting yourself to the glassware in Table 4.2, determine the proper combination of glassware to accomplish each dilution, and rank them in order of their most probable uncertainties.

13. Explain why changing all values in a data set by a constant amount will change $\overline{X}$ but has no effect on the standard deviation, *s*.

14. Obtain a sample of a metal, or other material, from your instructor and determine its density by one or both of the following methods:

    Method A: Determine the sample's mass with a balance. Calculate the sample's volume using appropriate linear dimensions.

    Method B: Determine the sample's mass with a balance. Calculate the sample's volume by measuring the amount of water it displaces by adding water to a graduated cylinder, reading the volume, adding the sample, and reading the new volume. The difference in volumes is equal to the sample's volume.

    Determine the density at least five times. (a) Report the mean, the standard deviation, and the 95% confidence interval for your results. (b) Find the accepted value for the metal's density and determine the absolute and relative error for your determination of the metal's density. (c) Use a propagation of uncertainty to determine the uncertainty for your method of analysis. Is the result of this calculation consistent with your experimental results? If not, suggest some possible reasons for this disagreement.

15. How many carbon atoms must a molecule have if the mean number of ^{13}C atoms per molecule is at least one? What percentage of such molecules will have no atoms of ^{13}C?

16. In Example 4.10 we determined the probability that a molecule of cholesterol, $C_{27}H_{44}O$, had no atoms of ^{13}C. (a) Calculate the probability that a molecule of cholesterol, has 1 atom of ^{13}C. (b) What is the probability that a molecule of cholesterol has two or more atoms of ^{13}C?

17. Berglund and Wichardt investigated the quantitative determination of Cr in high-alloy steels using a potentiometric titration of Cr(VI)[19]. Before the titration, samples of the steel were dissolved in acid and the chromium oxidized to Cr(VI) using peroxydisulfate. Shown here are the results (as %w/w Cr) for the analysis of a reference steel.

| 16.968 | 16.922 | 16.840 | 16.883 |
| 16.887 | 16.977 | 16.857 | 16.728 |

   Calculate the mean, the standard deviation, and the 95% confidence interval about the mean. What does this confidence interval mean?

18. Ketkar and co-workers developed an analytical method to determine trace levels of atmospheric gases.[20] An analysis of a sample that is 40.0 parts per thousand (ppt) 2-chloroethylsulfide gave the following results

43.3	34.8	31.9
37.8	34.4	31.9
42.1	33.6	35.3

   (a) Determine whether there is a significant difference between the experimental mean and the expected value at $\alpha = 0.05$. (b) As part of this study, a reagent blank was analyzed 12 times giving a mean of 0.16 ppt and a standard deviation of 1.20 ppt. What are the IUPAC detection limit, the limit of identification, and limit of quantitation for this method assuming $\alpha = 0.05$?

19. To test a spectrophotometer's accuracy a solution of 60.06 ppm $K_2Cr_2O_7$ in 5.0 mM H_2SO_4 is prepared and analyzed. This solution has an expected absorbance of 0.640 at 350.0 nm in a 1.0-cm cell when using 5.0 mM H_2SO_4 as a reagent blank. Several aliquots of the solution produce the following absorbance values.

   0.639   0.638   0.640   0.639   0.640   0.639   0.638

   Determine whether there is a significant difference between the experimental mean and the expected value at $\alpha = 0.01$.

20. Monna and co-workers used radioactive isotopes to date sediments from lakes and estuaries.[21] To verify this method they analyzed a ^{208}Po standard known to have an activity of 77.5 decays/min, obtaining the following results.

| 77.09 | 75.37 | 72.42 | 76.84 | 77.84 | 76.69 |
| 78.03 | 74.96 | 77.54 | 76.09 | 81.12 | 75.75 |

19 Berglund, B.; Wichardt, C. *Anal. Chim. Acta* **1990**, *236*, 399–410.
20 Ketkar, S. N.; Dulak, J. G.; Dheandhanou, S.; Fite, W. L. *Anal. Chim. Acta* **1991**, *245*, 267–270.
21 Monna, F.; Mathieu, D.; Marques, A. N.; Lancelot, J.; Bernat, M. *Anal. Chim. Acta* **1996**, *330*, 107–116.

Determine whether there is a significant difference between the mean and the expected value at $\alpha = 0.05$.

21. A 2.6540-g sample of an iron ore, which is 53.51% w/w Fe, is dissolved in a small portion of concentrated HCl and diluted to volume in a 250-mL volumetric flask. A spectrophotometric determination of the concentration of Fe in this solution yields results of 5840, 5770, 5650, and 5660 ppm. Determine whether there is a significant difference between the experimental mean and the expected value at $\alpha = 0.05$.

22. Horvat and co-workers used atomic absorption spectroscopy to determine the concentration of Hg in coal fly ash.[22] Of particular interest to the authors was developing an appropriate procedure for digesting samples and releasing the Hg for analysis. As part of their study they tested several reagents for digesting samples. Their results using HNO_3 and using a 1+3 mixture of HNO_3 and HCl are shown here. All concentrations are given as ppb Hg sample.

$$HNO_3: \quad 161 \quad 165 \quad 160 \quad 167 \quad 166$$
$$1{+}3 \ HNO_3{-}HCl: \quad 159 \quad 145 \quad 140 \quad 147 \quad 143 \quad 156$$

Determine whether there is a significant difference between these methods at $\alpha = 0.05$.

23. Lord Rayleigh, John William Strutt (1842-1919), was one of the most well known scientists of the late nineteenth and early twentieth centuries, publishing over 440 papers and receiving the Nobel Prize in 1904 for the discovery of argon. An important turning point in Rayleigh's discovery of Ar was his experimental measurements of the density of N_2. Rayleigh approached this experiment in two ways: first by taking atmospheric air and removing O_2 and H_2; and second, by chemically producing N_2 by decomposing nitrogen containing compounds (NO, N_2O, and NH_4NO_3) and again removing O_2 and H_2. The following table shows his results for the density of N_2, as published in *Proc. Roy. Soc.* **1894**, *LV*, 340 (publication 210); all values are the grams of gas at an equivalent volume, pressure, and temperature.[23]

Atmospheric Origin:	2.310 17	2.309 86	2.310 10	2.310 01
	2.310 24	2.310 10	2.310 28	
Chemical Origin:	2.301 43	2.298 90	2.298 16	2.301 82
	2.298 69	2.299 40	2.298 49	2.298 89

Explain why this data led Rayleigh to look for and to discover Ar.

22 Horvat, M.; Lupsina, V.; Pihlar, B. *Anal. Chim. Acta* **1991**, *243*, 71–79.
23 Larsen, R. D. *J. Chem. Educ.* **1990**, *67*, 925–928.

24. Gács and Ferraroli reported a method for monitoring the concentration of SO_2 in air.[24] They compared their method to the standard method by analyzing urban air samples collected from a single location. Samples were collected by drawing air through a collection solution for 6 min. Shown here is a summary of their results with SO_2 concentrations reported in $\mu L/m^3$.

standard method:	21.62	22.20	24.27	23.54
	24.25	23.09	21.02	
new method:	21.54	20.51	22.31	21.30
	24.62	25.72	21.54	

Using an appropriate statistical test, determine whether there is any significant difference between the standard method and the new method at $\alpha = 0.05$.

25. One way to check the accuracy of a spectrophotometer is to measure absorbances for a series of standard dichromate solutions obtained from the National Institute of Standards and Technology. Absorbances are measured at 257 nm and compared to the accepted values. The results obtained when testing a newly purchased spectrophotometer are shown here. Determine if the tested spectrophotometer is accurate at $\alpha = 0.05$.

Standard	Measured Absorbance	Expected Absorbance
1	0.2872	0.2871
2	0.5773	0.5760
3	0.8674	0.8677
4	1.1623	1.1608
5	1.4559	1.4565

26. Maskarinec and co-workers investigated the stability of volatile organics in environmental water samples.[25] Of particular interest was establishing the proper conditions to maintain the sample's integrity between its collection and its analysis. Two preservatives were investigated—ascorbic acid and sodium bisulfate—and maximum holding times were determined for a number of volatile organics and water matrices. The following table shows results for the holding time (in days) of nine organic compounds in surface water.

	Ascorbic Acid	Sodium Bisulfate
methylene chloride	77	62
carbon disulfide	23	54
trichloroethane	52	51

24 Gács, I.; Ferraroli, R. *Anal. Chim. Acta* **1992**, *269*, 177–185.
25 Maxkarinec, M. P.; Johnson, L. H.; Holladay, S. K.; Moody, R. L.; Bayne, C. K.; Jenkins, R. A. *Environ. Sci. Technol.* **1990**, *24*, 1665–1670.

	Ascorbic Acid	Sodium Bisulfate
benzene	62	42
1,1,2-trichloroethane	57	53
1,1,2,2-tetrachlorethane	33	85
tetrachloroethene	41	63
toluene	32	94
chlorobenzene	36	86

Determine whether there is a significant difference in the effectiveness of the two preservatives at $\alpha = 0.10$.

27. Using X-ray diffraction, Karstang and Kvalhein reported a new method to determine the weight percent of kaolinite in complex clay minerals using X-ray diffraction.[26] To test the method, nine samples containing known amounts of kaolinite were prepared and analyzed. The results (as % w/w kaolinite) are shown here.

Actual: 5.0   10.0   20.0   40.0   50.0   60.0   80.0   90.0   95.0
Found: 6.8   11.7   19.8   40.5   53.6   61.7   78.9   91.7   94.7

Evaluate the accuracy of the method at $\alpha = 0.05$.

28. Mizutani, Yabuki and Asai developed an electrochemical method for analyzing *l*-malate.[27] As part of their study they analyzed a series of beverages using both their method and a standard spectrophotometric procedure based on a clinical kit purchased from Boerhinger Scientific. The following table summarizes their results. All values are in ppm.

Sample	Electrode	Spectrophotometric
Apple juice 1	34.0	33.4
Apple juice 2	22.6	28.4
Apple juice 3	29.7	29.5
Apple juice 4	24.9	24.8
Grape juice 1	17.8	18.3
Grape juice 2	14.8	15.4
Mixed fruit juice 1	8.6	8.5
Mixed fruit juice 2	31.4	31.9
White wine 1	10.8	11.5
White wine 2	17.3	17.6
White wine 3	15.7	15.4
White wine 4	18.4	18.3

Determine whether there is a significant difference between the methods at $\alpha = 0.05$.

26 Karstang, T. V.; Kvalhein, O. M. *Anal. Chem.* **1991**, *63*, 767–772.
27 Mizutani, F.; Yabuki, S.; Asai, M. *Anal. Chim. Acta* **1991**, *245*,145–150.

29. Alexiev and colleagues describe an improved photometric method for determining Fe^{3+} based on its ability to catalyze the oxidation of sulphanilic acid by KIO_4.[28] As part of their study, the concentration of Fe^{3+} in human serum samples was determined by the improved method and the standard method. The results, with concentrations in μmol/L, are shown in the following table.

Sample	Improved Method	Standard Method
1	8.25	8.06
2	9.75	8.84
3	9.75	8.36
4	9.75	8.73
5	10.75	13.13
6	11.25	13.65
7	13.88	13.85
8	14.25	13.53

Determine whether there is a significant difference between the two methods at $\alpha = 0.05$.

30. Ten laboratories were asked to determine an analyte's concentration of in three standard test samples. Following are the results, in μg/mL.[29]

Laboratory	Sample 1	Sample 2	Sample 3
1	22.6	13.6	16.0
2	23.0	14.2	15.9
3	21.5	13.9	16.9
4	21.9	13.9	16.9
5	21.3	13.5	16.7
6	22.1	13.5	17.4
7	23.1	13.9	17.5
8	21.7	13.5	16.8
9	22.2	12.9	17.2
10	21.7	13.8	16.7

Determine if there are any potential outliers in Sample 1, Sample 2 or Sample 3. Use all three methods—Dixon's Q-test, Grubb's test, and Chauvenet's criterion—and compare the results to each other. For Dixon's Q-test and for the Grubb's test, use a significance level of $\alpha = 0.05$.

28 Alexiev, A.; Rubino, S.; Deyanova, M.; Stoyanova, A.; Sicilia, D.; Perez Bendito, D. *Anal. Chim. Acta*, **1994**, *295*, 211–219.

29 Data adapted from Steiner, E. H. "Planning and Analysis of Results of Collaborative Tests," in *Statistical Manual of the Association of Official Analytical Chemists*, Association of Official Analytical Chemists: Washington, D. C., 1975.

31. When copper metal and powdered sulfur are placed in a crucible and ignited, the product is a sulfide with an empirical formula of Cu_xS. The value of x is determined by weighing the Cu and the S before ignition and finding the mass of Cu_xS when the reaction is complete (any excess sulfur leaves as SO_2). The following table shows the Cu/S ratios from 62 such experiments.

1.764	1.838	1.865	1.866	1.872	1.877
1.890	1.891	1.891	1.897	1.899	1.900
1.906	1.908	1.910	1.911	1.916	1.919
1.920	1.922	1.927	1.931	1.935	1.936
1.936	1.937	1.939	1.939	1.940	1.941
1.941	1.942	1.943	1.948	1.953	1.955
1.957	1.957	1.957	1.959	1.962	1.963
1.963	1.963	1.966	1.968	1.969	1.973
1.975	1.976	1.977	1.981	1.981	1.988
1.993	1.993	1.995	1.995	1.995	2.017
2.029	2.042				

(a) Calculate the mean, the median, and the standard deviation for this data. (b) Construct a histogram for this data. From a visual inspection of your histogram, do the data appear normally distributed? (c) In a normally distributed population 68.26% of all members lie within the range $\mu \pm 1\sigma$. What percentage of the data lies within the range $\overline{X} \pm 1s$? Does this support your answer to the previous question? (d) Assuming that $\overline{X}$ and s^2 are good approximations for μ and for σ^2, what percentage of all experimentally determined Cu/S ratios should be greater than 2? How does this compare with the experimental data? Does this support your conclusion about whether the data is normally distributed? (e) It has been reported that this method of preparing copper sulfide results in a non-stoichiometric compound with a Cu/S ratio of less than 2. Determine if the mean value for this data is significantly less than 2 at a significance level of $\alpha = 0.01$.

See Blanchnik, R.; Müller, A. "The Formation of Cu_2S From the Elements I. Copper Used in Form of Powders," *Thermochim. Acta*, **2000**, *361*, 31-52 for a discussion of some of the factors affecting the formation of non-stoichiometric copper sulfide.

32. Real-time quantitative PCR is an analytical method for determining trace amounts of DNA. During the analysis, each cycle doubles the amount of DNA. A probe species that fluoresces in the presence of DNA is added to the reaction mixture and the increase in fluorescence is monitored during the cycling. The cycle threshold, C_p, is the cycle when the fluorescence exceeds a threshold value. The data in the following table shows C_t values for three samples using real-time quantitative PCR.[30] Each sample was analyzed 18 times.

30 Burns, M. J.; Nixon, G. J.; Foy, C. A.; Harris, N. *BMC Biotechnol.* **2005**, *5:31* (open access publication).

Sample X		Sample Y		Sample Z	
24.24	25.14	24.41	28.06	22.97	23.43
23.97	24.57	27.21	27.77	22.93	23.66
24.44	24.49	27.02	28.74	22.95	28.79
24.79	24.68	26.81	28.35	23.12	23.77
23.92	24.45	26.64	28.80	23.59	23.98
24.53	24.48	27.63	27.99	23.37	23.56
24.95	24.30	28.42	28.21	24.17	22.80
24.76	24.60	25.16	28.00	23.48	23.29
25.18	24.57	28.53	28.21	23.80	23.86

Examine this data and write a brief report on your conclusions. Issues you may wish to address include the presence of outliers in the samples, a summary of the descriptive statistics for each sample, and any evidence for a difference between the samples.

4L   Solutions to Practice Exercises

Practice Exercise 4.1

Mean: To find the mean we sum the individual measurements and divide by the number of measurements. The sum of the 10 concentrations is 1405. Dividing the sum by 10, gives the mean as 140.5, or 1.40×10^2 mmol/L.

Median: To find the mean we arrange the 10 measurements from the smallest concentration to the largest concentration; thus

118 132 137 140 141 143 143 145 149 157

The median for a data set with 10 members is the average of the fifth and sixth values; thus, the median is (141+143)/2, or 142 mmol/L.

Range: The range is the difference between the largest value and the smallest value; thus, the range is $157 - 118 = 39$ mmol/L.

Standard Deviation: To calculate the standard deviation we first calculate the difference between each measurement and the mean value (140.5), square the resulting differences, and add them together. The differences are

–0.5 2.5 0.5 –3.5 –8.5 16.5 2.5 8.5 –22.5 4.5

and the squared differences are

0.25 6.25 0.25 12.25 72.25 272.25 6.25 72.25 506.25 20.25

The total sum of squares, which is the numerator of equation 4.1, is 968.50. The standard deviation is

$$s = \sqrt{\frac{968.50}{10 - 1}} = 10.37 \approx 10.4$$

Variance: The variance is the square of the standard deviation, or 108.

Click here to return to the chapter.

Practice Exercise 4.2

The first step is to determine the concentration of Cu^{2+} in the final solution. The mass of copper is

$$74.2991 \text{ g} - 73.3216 \text{ g} = 0.9775 \text{ g Cu}$$

The 10 mL of HNO_3 used to dissolve the copper does not factor into our calculation. The concentration of Cu^{2+} is

$$\frac{0.9775 \text{ g Cu}}{0.5000 \text{ L}} \times \frac{1.000 \text{ mL}}{250.0 \text{ mL}} \times \frac{1000 \text{ mg}}{\text{g}} = 7.820 \text{ mg } Cu^{2+}/\text{L}$$

Having found the concentration of Cu^{2+}, we continue with the propagation of uncertainty. The absolute uncertainty in the mass of Cu wire is

$$u_{\text{g Cu}} = \sqrt{(0.0001)^2 + (0.0001)^2} = 0.00014 \text{ g}$$

The relative uncertainty in the concentration of Cu^{2+} is

$$\frac{u_{\text{mg/L}}}{7.820 \text{ mg/L}} =$$
$$\sqrt{\left(\frac{0.00014}{0.9775}\right)^2 + \left(\frac{0.20}{500.0}\right)^2 + \left(\frac{0.006}{1.000}\right)^2 + \left(\frac{0.12}{250.0}\right)^2} = 0.00603$$

Solving for $u_{\text{mg/L}}$ gives the uncertainty as 0.0472. The concentration and uncertainty for Cu^{2+} is 7.820 mg/L $\pm$ 0.047 mg/L.

Click here to return to the chapter.

Practice Exercise 4.3

The first step is to calculate the absorbance, which is

$$A = -\log \frac{P}{P_o} = -\log \frac{1.50 \times 10^2}{3.80 \times 10^2} = 0.4037 \approx 0.404$$

Having found the absorbance, we continue with the propagation of uncertainty. First, we find the uncertainty for the ratio P/P_o.

$$\frac{u_{P/P_o}}{P/P_o} = \sqrt{\left(\frac{15}{3.80 \times 10^2}\right)^2 + \left(\frac{15}{1.50 \times 10^2}\right)^2} = 0.1075$$

Finally, from Table 4.10 the uncertainty in the absorbance is

$$u_A = 0.4343 \times \frac{u_{P/P_o}}{P/P_o} = (0.4343) \times (0.1075) = 4.669 \times 10^{-2}$$

The absorbance and uncertainty is 0.404 $\pm$ 0.005 absorbance units.

Click here to return to the chapter.

Practice Exercise 4.4

An uncertainty of 0.8% is a relative uncertainty in the concentration of 0.008; thus

$$0.008 = \sqrt{\left(\frac{0.028}{23.41}\right)^2 + \left(\frac{u_{k_A}}{0.186}\right)^2}$$

Squaring both sides of the equation gives

$$6.4 \times 10^{-5} = \left(\frac{0.028}{23.41}\right)^2 + \left(\frac{u_{k_A}}{0.186}\right)^2$$

Solving for u_{k_A} gives its value as 1.47×10^{-3} or ± 0.0015 ppm^{-1}.

Click here to return to the chapter.

Practice Exercise 4.5

To find the percentage of tablets that contain less than 245 mg of aspirin we first calculate the deviation, z,

$$z = \frac{245 - 250}{5} = -1.00$$

and then look up the corresponding probability in Appendix 3A, obtaining a value of 15.87%. To find the percentage of tablets that contain less than 240 mg of aspirin we find that

$$z = \frac{240 - 250}{5} = -2.00$$

which corresponds to 2.28%. The percentage of tablets containing between 240 and 245 mg of aspiring is $15.87\% - 2.28\% = 13.59\%$.

Click here to return to the chapter.

Practice Exercise 4.6

The mean is 249.9 mg aspirin/tablet for this sample of seven tablets. For a 95% confidence interval the value of z is 1.96, which makes the confidence interval

$$249.9 \pm \frac{1.96 \times 5}{\sqrt{7}} = 249.9 \pm 3.7 \approx 250 \text{ mg} \pm 4 \text{ mg}$$

Click here to return to the chapter.

Practice Exercise 4.7

With 100 pennies, we have 99 degrees of freedom for the mean. Although Table 4.15 does not include a value for $t(0.05, 99)$, we can approximate its value by using the values for $t(0.05, 60)$ and $t(0.05, 100)$ and by assuming a linear change in its value.

$$t(0.05, 99) = t(0.05, 60) - \frac{39}{40}\{t(0.05, 60) - t(0.05, 100)\}$$

$$t(0.05, 99) = 2.000 - \frac{39}{40}\{2.000 - 1.984\} = 1.9844$$

The 95% confidence interval for the pennies is

$$3.095 \pm \frac{1.9844 \times 0.0346}{\sqrt{100}} = 3.095 \text{ g} \pm 0.007 \text{g}$$

From Example 4.15, the 95% confidence intervals for the two samples in Table 4.11 are 3.117 g $\pm$ 0.047 g and 3.081 g $\pm$ 0.046 g. As expected, the confidence interval for the sample of 100 pennies is much smaller than that for the two smaller samples of pennies. Note, as well, that the confidence interval for the larger sample fits within the confidence intervals for the two smaller samples.

Click here to return to the chapter.

Practice Exercise 4.8

The null hypothesis is $H_0: \overline{X} = \mu$ and the alternative hypothesis is $H_A: \overline{X} \neq \mu$. The mean and the standard deviation for the data are 99.26% and 2.35%, respectively. The value for t_{exp} is

$$t_{exp} = \frac{|100.0 - 99.26| \sqrt{7}}{2.35} = 0.833$$

and the critical value for $t(0.05, 6)$ is 0.836. Because t_{exp} is less than $t(0.05, 6)$ we retain the null hypothesis and have no evidence for a significant difference between $\overline{X}$ and μ.

Click here to return to the chapter.

Practice Exercise 4.9

The standard deviations are 6.451 mg for Lot 1 and 7.849 mg for Lot 2. The null and alternative hypotheses are

$$H_0: s_{Lot 1}^2 = s_{Lot 2}^2 \qquad H_A: s_{Lot 1}^2 \neq s_{Lot 2}^2$$

and the value of F_{exp} is

$$F_{exp} = \frac{(7.849)^2}{(6.451)^2} = 1.480$$

The critical value for $F(0.05, 5, 6)$ is 5.988. Because $F_{exp} < F(0.05, 5, 6)$, we retain the null hypothesis. There is no evidence at $\alpha = 0.05$ to suggest that the difference in the variances is significant.

Click here to return to the chapter.

Practice Exercise 4.10

To compare the means for the two lots, we use an unpaired t-test of the null hypothesis $H_0: \overline{X}_{Lot 1} = \overline{X}_{Lot 2}$ and the alternative hypothesis $H_A: \overline{X}_{Lot 1} \neq \overline{X}_{Lot 2}$. Because there is no evidence to suggest a difference in the variances (see Practice Exercise 4.9) we pool the standard deviations, obtaining an s_{pool} of

$$s_{pool} = \sqrt{\frac{(7-1)(6.451)^2 + (6-1)(7.849)^2}{7+6-2}} = 7.121$$

The means for the two samples are 249.57 mg for Lot 1 and 249.00 mg for Lot 2. The value for t_{exp} is

$$t_{exp} = \frac{|249.57 - 249.00|}{7.121} \times \sqrt{\frac{7 \times 6}{7+6}} = 0.1439$$

The critical value for $t(0.05, 11)$ is 2.204. Because t_{exp} is less than $t(0.05, 11)$, we retain the null hypothesis and find no evidence at $\alpha = 0.05$ that there is a significant difference between the means for the two lots of aspirin tablets.

Click here to return to the chapter.

Practice Exercise 4.11

Treating as Unpaired Data: The mean and the standard deviation for the concentration of Zn^{2+} at the air-water interface are 0.5178 mg/L and 0.1732 mg/L, respectively, and the values for the sediment-water interface are 0.4445 mg/L and 0.1418 mg/L, respectively. An *F*-test of the variances gives an F_{exp} of 1.493 and an $F(0.05, 5, 5)$ of 7.146. Because F_{exp} is smaller than $F(0.05, 5, 5)$, we have no evidence at $\alpha = 0.05$ to suggest that the difference in variances is significant. Pooling the standard deviations gives an s_{pool} of 0.1582 mg/L. An unpaired *t*-test gives t_{exp} as 0.8025. Because t_{exp} is smaller than $t(0.05, 11)$, which is 2.204, we have no evidence that there is a difference in the concentration of Zn^{2+} between the two interfaces.

Treating as Paired Data: To treat as paired data we need to calculate the difference, d_i, between the concentration of Zn^{2+} at the air-water interface and at the sediment-water interface for each location, where

$$d_i = \left([Zn^{2+}]_{air-water}\right)_i - \left([Zn^{2+}]_{sed-water}\right)_i$$

Location	1	2	3	4	5	6
d_i (mg/L)	0.015	0.028	0.067	0.121	0.102	0.107

The mean difference is 0.07333 mg/L with a standard deviation of 0.0441 mg/L. The null hypothesis and the alternative hypothesis are

$$H_0: \overline{d} = 0 \qquad H_A: \overline{d} \neq 0$$

and the value of t_{exp} is

$$t_{exp} = \frac{|0.07333|\sqrt{6}}{0.04410} = 4.073$$

Because t_{exp} is greater than $t(0.05, 5)$, which is 2.571, we reject the null hypothesis and accept the alternative hypothesis that there is a significant difference in the concentration of Zn^{2+} between the air-water interface and the sediment-water interface.

The difference in the concentration of Zn^{2+} between locations is much larger than the difference in the concentration of Zn^{2+} between the interfaces. Because out interest is in studying the difference between the interfaces, the larger standard deviation when treating the data as unpaired increases the probability of incorrectly retaining the null hypothesis, a type 2 error.

Click <u>here</u> to return to the chapter.

Practice Exercise 4.12

You will find small differences between the values you obtain using Excel's built in functions and the worked solutions in the chapter. These differences arise because Excel does not round off the results of intermediate calculations.

Click <u>here</u> to return to the chapter.

Practice Exercise 4.13

Shown here are copies of R sessions for each problem. You will find small differences between the values given here for t_{exp} and for F_{exp} and those values shown with the worked solutions in the chapter. These differences arise because R does not round off the results of intermediate calculations.

Example 4.20

```
> AnalystA = c(86.82, 87.04, 86.93, 87.01, 86.20, 87.00)
> AnalystB = c(81.01, 86.15, 81.73, 83.19, 80.27, 83.94)
> var.test(AnalystB, AnalystA)

        F test to compare two variances

data:  AnalystB and AnalystA
F = 45.6358, num df = 5, denom df = 5, p-value = 0.0007148
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  6.385863 326.130970
sample estimates:
ratio of variances
       45.63582

> t.test(AnalystA, AnalystB, var.equal=FALSE)

        Welch Two Sample t-test

data:  AnalystA and AnalystB
t = 4.6147, df = 5.219, p-value = 0.005177
alternative hypothesis: true difference in means is not equal to 0
```

95 percent confidence interval:
 1.852919 6.383748
sample estimates:
mean of x mean of y
 86.83333 82.71500

Example 4.21

> micro = c(129.5, 89.6, 76.6, 52.2, 110.8, 50.4, 72.4, 141.4, 75.0, 34.1, 60.3)
> elect = c(132.3, 91.0, 73.6, 58.2, 104.2, 49.9, 82.1, 154.1, 73.4, 38.1, 60.1)
> t.test(micro,elect,paired=TRUE)

        Paired t-test

data: micro and elect
t = -1.3225, df = 10, p-value = 0.2155
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.028684 1.537775
sample estimates:
mean of the differences
        -2.245455

Click here to return to the chapter.

Practice Exercise 4.14

Because we are selecting a random sample of 100 members from a uniform distribution, you will see subtle differences between your plots and the plots shown as part of this answer. Here is a record of my R session and the resulting plots.

> data = runif(100, min = 0, max = 0)
> data
 [1] 18.928795 80.423589 39.399693 23.757624 30.088554
 [6] 76.622174 36.487084 62.186771 81.115515 15.726404
[11] 85.765317 53.994179 7.919424 10.125832 93.153308
[16] 38.079322 70.268597 49.879331 73.115203 99.329723
[21] 48.203305 33.093579 73.410984 75.128703 98.682127
[26] 11.433861 53.337359 81.705906 95.444703 96.843476
[31] 68.251721 40.567993 32.761695 74.635385 70.914957
[36] 96.054750 28.448719 88.580214 95.059215 20.316015
[41] 9.828515 44.172774 99.648405 85.593858 82.745774
[46] 54.963426 65.563743 87.820985 17.791443 26.417481
[51] 72.832037 5.518637 58.231329 10.213343 40.581266
[56] 6.584000 81.261052 48.534478 51.830513 17.214508
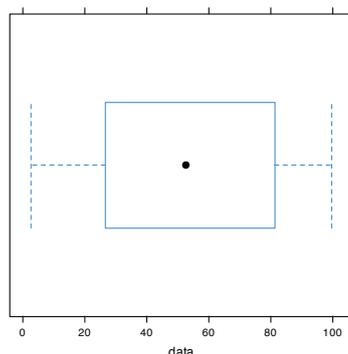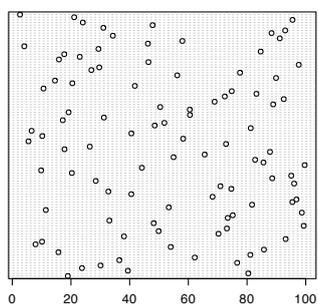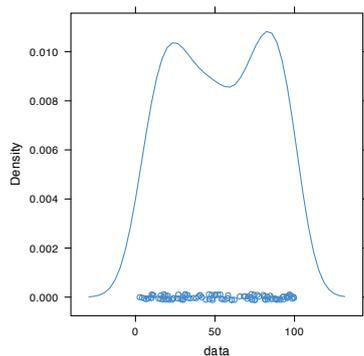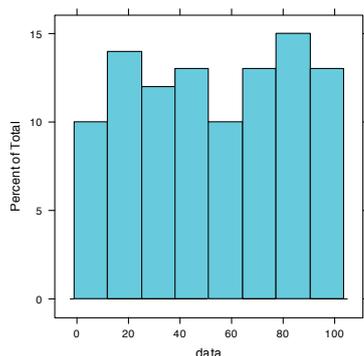
[61] 31.232099 60.545307 19.197450 60.485374 50.414960
[66] 88.908862 68.939084 92.515781 72.414388 83.195206
[71] 74.783176 10.643619 41.775788 20.464247 14.547841
[76] 89.887518 56.217573 77.606742 26.956787 29.641171
[81] 97.624246 46.406271 15.906540 23.007485 17.715668
[86] 84.652814 29.379712 4.093279 46.213753 57.963604
[91] 91.160366 34.278918 88.352789 93.004412 31.055807
[96] 47.822329 24.052306 95.498610 21.089686 2.629948
> histogram(data, type = "percent")
> densityplot(data)
> dotchart(data)
> bwplot(data)

Figure 4.26 shows the four plots. The histogram divides the data into eight bins, each of which contains between 10 and 15 members. As we expect for a uniform distribution, the histogram's overall pattern suggests that each outcome is equally probable. In interpreting the kernel density plot, it is important to remember that it treats each data point as if it is from a normally distributed population (even though, in this case, the underlying population is uniform). Although the plot appears to suggest that there are two normally distributed populations, the individual results shown at the bottom of the plot provide further evidence for a uniform distribution. The dot chart shows no trend along the *y*-axis, which indicates that the individual members of this sample were drawn at random from the population. The distribution along the *x*-axis also shows no pattern, as expected for a uniform distribution, Finally, the box plot shows no evidence of outliers.

Click here to return to the chapter.

Figure 4.26 Plots generated using R to solve Practice Exercise 4.13.

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

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

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 that contain known amounts of analyte. The accuracy of a standardization, therefore, depends on the quality of the reagents and the glassware we use 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 used to reach the equivalence point. The accuracy of a titrimetric analysis, therefore, is never better than the accuracy with which we know the titrant's concentration.

5A.1   Primary and Secondary Standards

There are two categories of analytical standards: primary standards and secondary standards. A PRIMARY STANDARD is a reagent that we can use to 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 $K_2Cr_2O_7$ in 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 it is difficult to establishing accurately 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 is determined relative to a primary standard. Lists of acceptable primary standards are available.[2] Appendix 8 provides examples of some common primary standards.

5A.2   Other Reagents

Preparing a standard often requires additional reagents that are not primary standards or secondary standards, such as a suitable solvent or reagents needed 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 that conform to standards set by the American Chemical Society are 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 can correct for contributions to S_{total} from reagents used in an

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

NaOH is one 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, we titrate it against a primary standard weak acid, such as potassium hydrogen phthalate, $KHC_8H_4O_4$.

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.

analysis by including an appropriate blank determination in the analytical procedure.

5A.3   Preparing a Standard Solution

It often is 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 a larger range of concentrations, particularly a range that extends 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 that solution 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 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.
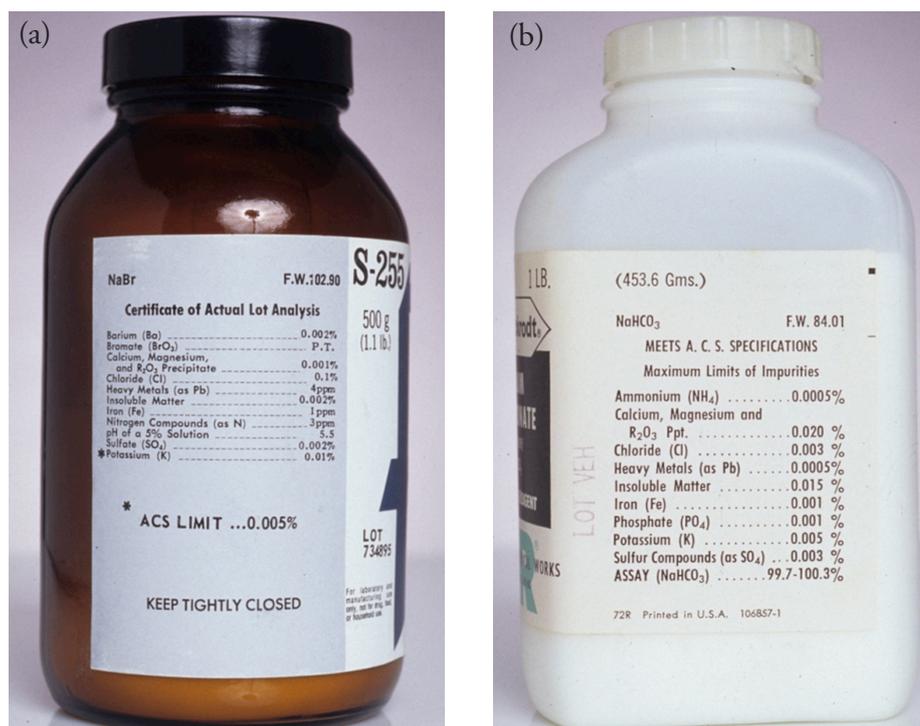
Figure 5.1 Two examples of packaging labels for reagent grade chemicals. The label in (a) provides the manufacturer's assay for the reagent, NaBr. Note that potassium is flagged with an asterisk (*) because its assay exceeds the limit established by the American Chemical Society (ACS). The label in (b) does not provide an assay for impurities; however it indicates that the reagent meets ACS specifications by providing the maximum limits for impurities. An assay for the reagent, $NaHCO_3$, is provided.

5B   Calibrating the Signal (S_{total})

The accuracy with which we determine 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 that might affect the signal, we first calibrate our equipment and instrumentation 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 that focus 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}.

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 calibrated properly.

5C   Determining the Sensitivity (k_A)

To standardize an analytical method we also must determine the analyte's sensitivity, 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 is possible to derive the value of k_A for any analytical method if we understand fully all the chemical reactions and physical processes responsible for the signal. Unfortunately, such calculations are not feasible if we lack a sufficiently developed theoretical model of the physical processes or if the chemical reaction's evince non-ideal behavior. In such situations we must determine the value of k_A by analyzing one or more standard solutions, each of which contains a known amount of analyte. In this section we consider several approaches for determining the value of k_A. For simplicity we assume that S_{reag} is 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.

---

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

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 when measuring mass. See Appendix 9 for a discussion of correcting for the buoyancy of air.

Be sure to read and follow carefully the calibration instructions provided with any instrument you use.

$$S_A = k_A n_A \qquad 5.3$$

$$S_A = k_A C_A \qquad 5.4$$

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

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

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

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

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

A single-point standardization is the least desirable method for standardizing a method. There are 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 based on a single concentration of analyte. To extend this value of k_A to other concentrations of analyte requires that we 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 leads to a determinate error in C_A if k_A becomes smaller at higher concentrations of analyte. Despite these limitations, single-point standardizations find routine use when the expected range for the analyte's concentrations is small. Under these conditions it often is safe

---

5   Cardone, M. J.; Palmero, P. J.; Sybrandt, L. B. *Anal. Chem.* **1980**, *52*, 1187–1191.
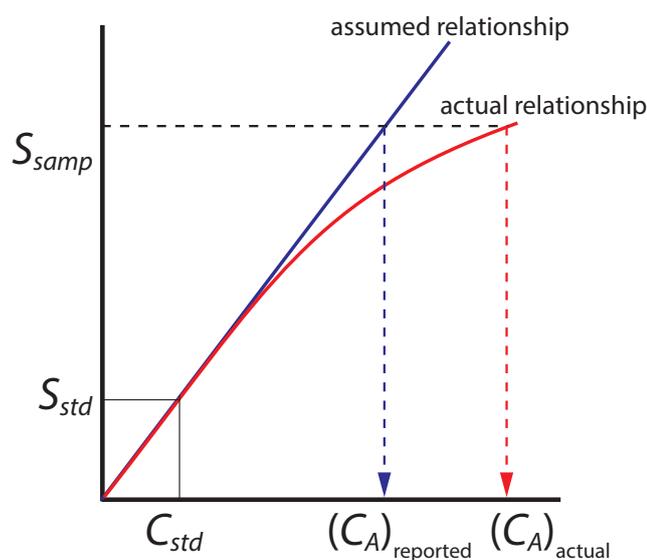
**Figure 5.2** Example showing how a single-point standardization leads to a determinate error in an analyte's reported concentration if we incorrectly assume that k_A is constant. The **assumed** relationship between S_{samp} and C_A is based on a single standard and is a straight-line; the **actual** relationship between S_{samp} and C_A becomes curved for larger concentrations of analyte.

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 better way to standardize a method is to prepare a series of standards, each of which contains a different concentration of analyte. 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 C_{std} is called a CALIBRATION CURVE. The exact standardization, or calibration relationship, is determined by an appropriate curve-fitting algorithm.

There are two advantages to a multiple-point standardization. First, although a determinate error in one standard introduces a determinate error, its effect is minimized by the remaining standards. Second, because we measure the signal for several concentrations of analyte, we no longer must assume k_A is independent of the analyte's concentration. Instead, we can construct a calibration curve similar to the "actual relationship" in Figure 5.2.

5C.2   External Standards

The most common method of standardization uses one or more EXTERNAL STANDARDS, each of which contains a known concentration of analyte. We call these standards "external" because they are prepared and analyzed separate from the samples.

SINGLE EXTERNAL STANDARD

With a single external standard we determine k_A using equation 5.5 and then calculate the concentration of analyte, C_A, 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 for which the 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 using the data for the single external standard.

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

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

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

MULTIPLE EXTERNAL STANDARDS

Figure 5.3 shows a typical multiple-point external standardization. The volumetric flask on the left contains 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 because the method's sensitivity remains constant throughout the analyte's concentration range. When the calibration curve is not a straight-line, the 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 it decreases continuously for higher concentrations of analyte. The value of k_A at any point along the calibration curve in Figure 5.2 is the slope at that point. In either case, a calibration curve allows to relate 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?
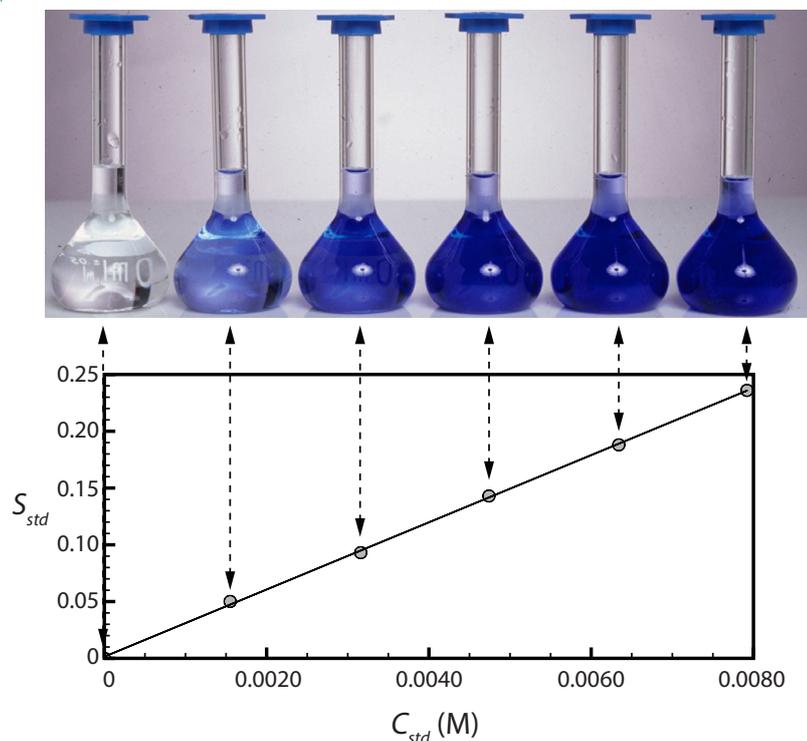
Figure 5.3 The photo at the top of the figure shows a reagent blank (far left) and a set of five external standards for Cu^{2+} with concentrations that increase from left-to-right. Shown below the external standards is the resulting normal calibration curve. The absorbance of each standard, S_{std}, is shown by the filled circles.

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 our 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.

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

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 present in the external standard's matrix, which usually is a much simpler matrix than that of our samples. When we use an external standardization we assume 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 an analyte in the sample's matrix and in the standard's matrix. In this case, using the calibration curve for the external standards leads to a negative determinate error in analyte's reported concentration. If we expect that matrix effects are important, then we try to match the standard's matrix to that of the sample, a process 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.
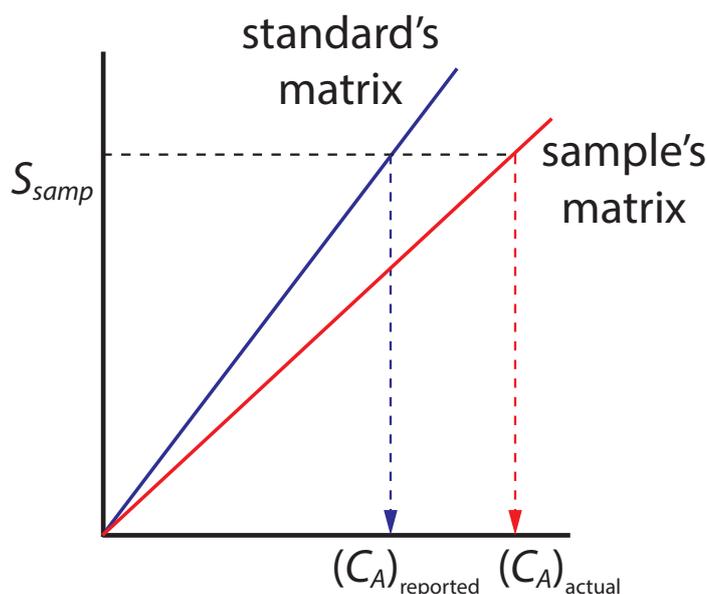
Figure 5.4 Calibration curves for an analyte in the standard's matrix and in the sample's matrix. If the matrix affects the value of k_A, as is the case here, then we introduce a proportional determinate error into our analysis if we use a normal calibration curve.

5C.3   Standard Additions

We can avoid the complication of matching the matrix of the standards to the matrix of the sample if we carry out 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 we dilute 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.

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

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

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

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

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

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

add V_o of C_A add V_{std} of C_{std}

dilute to V_f

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.

Concentration of Analyte

$$C_A \times \frac{V_o}{V_f} \qquad C_A \times \frac{V_o}{V_f} + C_{std} \times \frac{V_{std}}{V_f}$$
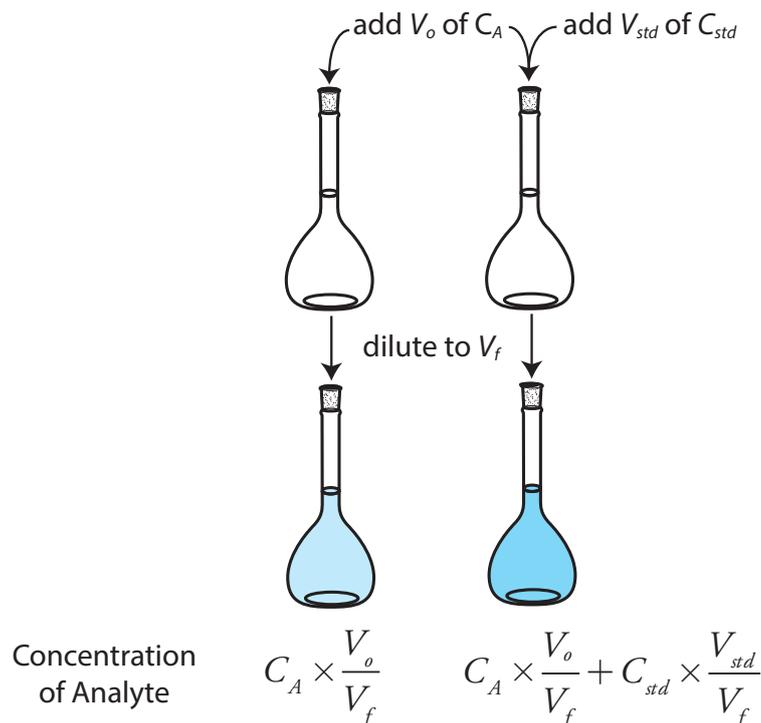
Example 5.3

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.

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

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

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

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

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

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

It also is possible to add the 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

$$S_{samp} = k_A C_A$$

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

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

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.

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

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

$V_o + V_{std} = 5.000\ mL + 5.00 \times 10^{-3}\ mL$
$\qquad\qquad = 5.005\ mL$

add V_{std} of C_{std}

Concentration of Analyte $\qquad C_A \qquad\qquad C_A \dfrac{V_o}{V_o + V_{std}} + C_{std} \dfrac{V_{std}}{V_o + V_{std}}$
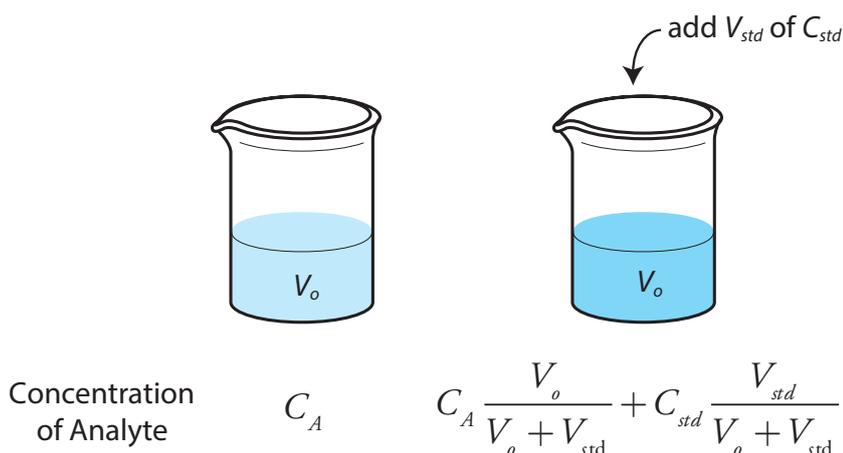
Figure 5.6 Illustration showing an alternative form of the method of standard additions. In this case we add the spike of external standard directly to the sample without any further adjust in the volume.

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

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

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

MULTIPLE STANDARD ADDITIONS

We can adapt a single-point standard addition into a multiple-point standard addition by preparing a series of samples that contain 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 $-C_A V_o/C_{std}$.

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

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

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

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.

(a)

$y\text{-intercept} = \dfrac{k_A C_A V_o}{V_f}$

$\text{slope} = \dfrac{k_A C_{std}}{V_f}$

V_{std} (mL)

$x\text{-intercept} = \dfrac{-C_A V_o}{C_{std}}$

(b)

$y\text{-intercept} = \dfrac{k_A C_A V_o}{V_f}$

$\text{slope} = k_A$

$x\text{-intercept} = \dfrac{-C_A V_o}{V_f}$

$C_{std} \times \dfrac{V_{std}}{V_f}$ (mg/L)
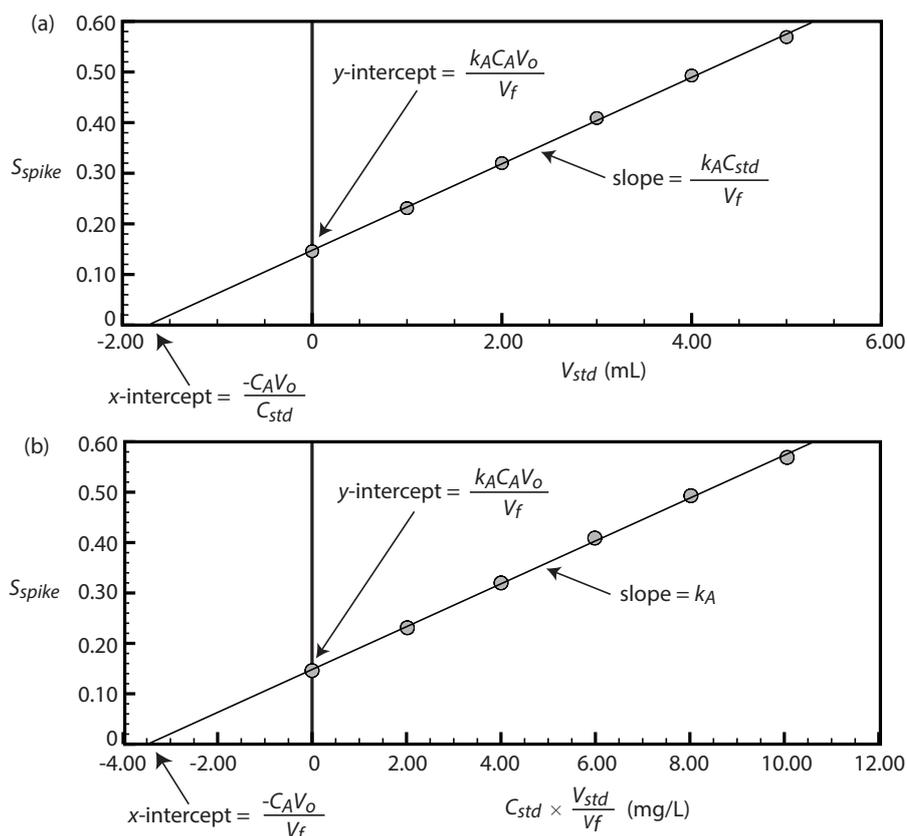
Figure 5.7 Shown at the top of the figure is a set of six standard additions for the determination of Mn^{2+}. The flask on the left is a 25.00 mL sample diluted to 50.00 mL with water. The remaining flasks contain 25.00 mL of sample and, from left-to-right, 1.00, 2.00, 3.00, 4.00, and 5.00 mL spikes of an external standard that is 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 5.6

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-intercept'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 with water 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 five-point calibration curve. For a normal calibration curve you need to analyze only 15 solutions (five standards and ten samples). If you use the method of standard additions, however, you must analyze 50 solutions (each of the ten samples is analyzed five times, once before spiking and after each of four 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 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 will increase if 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 remain identical. In effect, we can ignore evaporation if the samples and the 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 are no longer 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 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 the internal standard, respectively. 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 determine independently values for either k_A or k_{IS}.

SINGLE INTERNAL STANDARD

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

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

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 that is 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 spiked with the same concentration of Cu^{2+} gives a signal ratio, $(S_A/S_{IS})_{samp}$, of 1.80. What is 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

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

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

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

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 of which contains 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.

$$C_A = \frac{\left(\frac{S_A}{S_{IS}}\right)_{samp} + 0.006}{2.11\ \text{ppb}^{-1}} = \frac{2.80 + 0.006}{2.11\ \text{ppb}^{-1}} = 1.33\ \text{Pb}^{2+}$$

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 we prepare 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 that contains a known concentration of analyte. Using this value of k_A and our sample's signal, 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 for the five trials. 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 using the data in Table 5.1? In a single-point standardization we assume that the 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 constant determinate error. Table 5.2 demonstrates how an uncorrected constant error affects our determination

You might wonder if it is possible to include an internal standard in the method of standard additions to correct for both matrix effects and uncontrolled variations between samples; well, the answer is yes as described in the paper "Standard Dilution Analysis," the full reference for which is Jones, W. B.; Donati, G. L.; Calloway, C. P.; Jones, B. T. *Anal. Chem.* **2015**, *87*, 2321-2327.

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

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

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

of k_A. The first three columns show the concentration of analyte in a set of 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 each apparent k_A is 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 determining the best equation for the calibration curve is called linear regression.

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

$$y = \beta_0 + \beta_1 x \qquad\qquad 5.14$$

where y is the analyte's 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 method for completing the linear regression for equation 5.14 makes three assumptions:
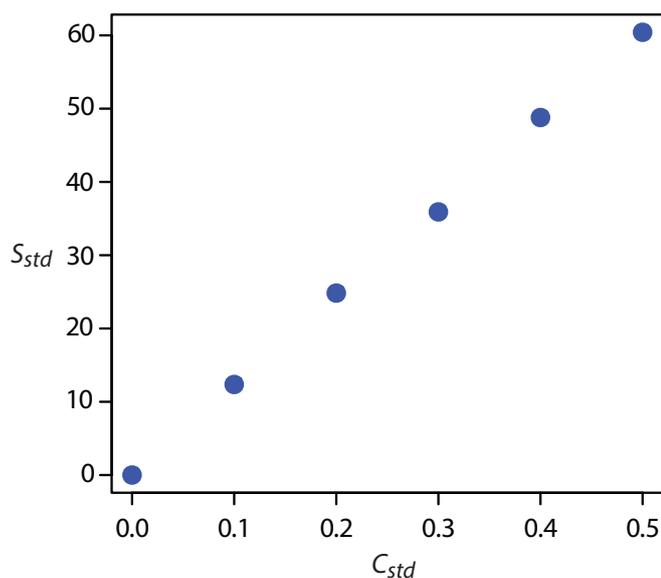
Figure 5.8 Normal calibration curve data for the hypothetical multiple-point external standardization in Table 5.1.

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

(2) that indeterminate errors that affect 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 our estimate of the slope and the y-intercept. For this reason the result is considered an UNWEIGHTED LINEAR REGRESSION.

The second assumption generally is 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 you accept the results of a linear regression. In particular the first assumption always is suspect because there certainly is some indeterminate error in the measurement of x. When we prepare a calibration curve, however, it is not unusual to find that the uncertainty in the signal, S_{std}, is significantly larger than the uncertainty in the analyte's concentration, C_{std}. In such circumstances the first assumption is usually reasonable.

HOW A LINEAR REGRESSION WORKS

To understand the logic of a 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 fit 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

$$\widehat{y} = b_0 + b_1 x \qquad\qquad 5.15$$
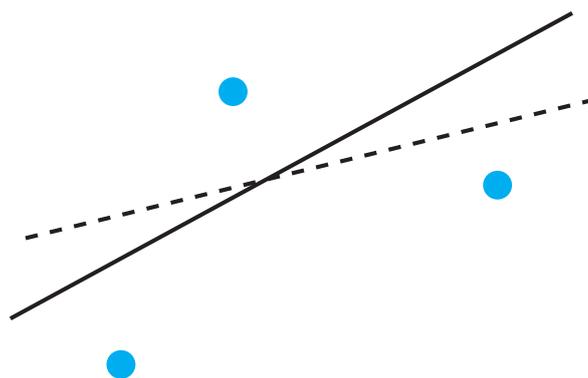
Figure 5.9 Illustration showing three data points and two possible straight-lines that might explain the data. The goal of a linear regression is to find the mathematical model, in this case a straight-line, that best explains the data.

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

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

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

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

$$R = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad 5.16$$

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 a positive residual error from canceling out a negative residual error. 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.
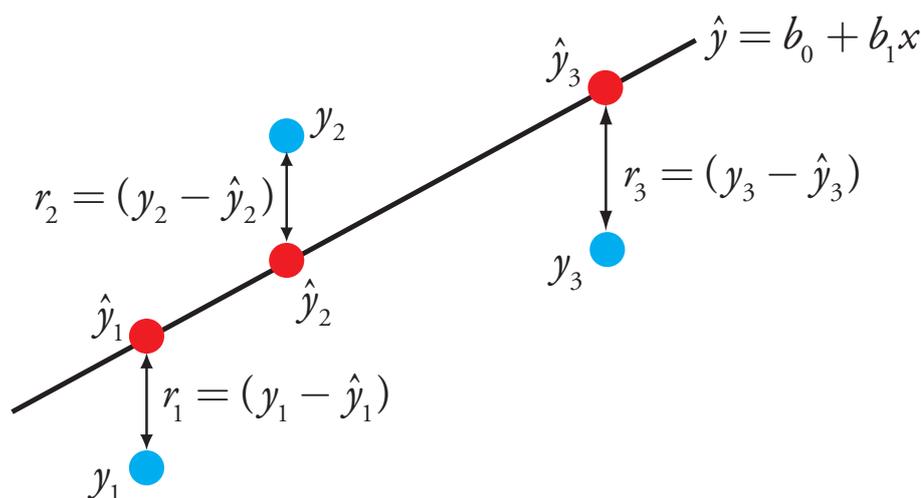
Figure 5.10 Illustration showing the evaluation of a linear regression in which we assume that all uncertainty is the result of indeterminate errors in y. The points in **blue**, y_i, are the original data and the points in **red**, $\hat{y}_i$, are the predicted values from the regression equation, $\hat{y} = b_0 + b_1 x$. 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

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad 5.17$$

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

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n} \qquad 5.18$$

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

$$\sum_{i=1}^{n} x_i \qquad \sum_{i=1}^{n} y_i \qquad \sum_{i=1}^{n} x_i y_i \qquad \sum_{i=1}^{n} x_i^2$$

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.

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

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

$$\sum_{i=1}^{n} x_i = 1.500 \quad \sum_{i=1}^{n} y_i = 182.31 \quad \sum_{i=1}^{n} x_i y_i = 66.701 \quad \sum_{i=1}^{n} x_i^2 = 0.550$$

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

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}.

---

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

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

$$b_1 = \frac{182.31 - (120.706 \times 1.500)}{6} = 0.209 \approx 0.21$$

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

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

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 indeterminate errors in the 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

$$s_r = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - 2}} \qquad 5.19$$

where y_i is the ith 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 $\hat{y}_i$.

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)?

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})}{n - 1}}$$
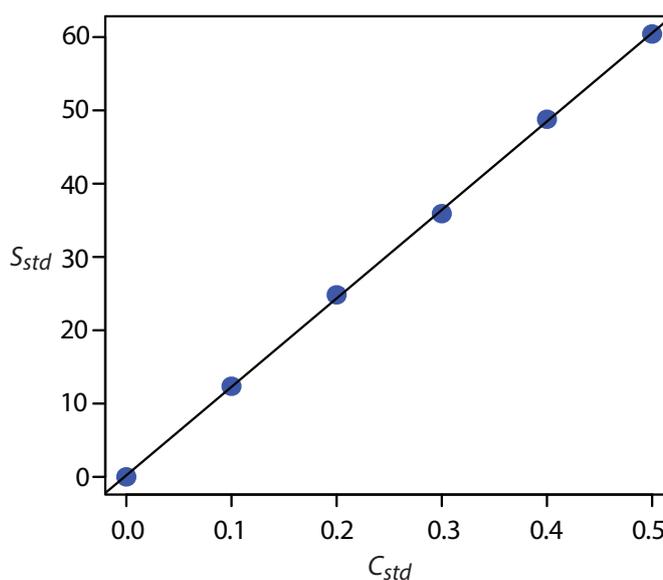
Figure 5.11 Calibration curve for the data in Table 5.1 and Example 5.9.

A more useful representation of the uncertainty in our regression analysis 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.

$$s_{b_1} = \sqrt{\frac{ns_r^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = \sqrt{\frac{s_r^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \qquad 5.20$$

$$s_{b_0} = \sqrt{\frac{s_r^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = \sqrt{\frac{s_r^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} (x_i - \overline{x})^2}} \qquad 5.21$$

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

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

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

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.

> You might contrast equation 5.22 and equation 5.23 with equation 4.12
>
> $$\mu = \overline{X} \pm \frac{ts}{\sqrt{n}}$$
>
> for the confidence interval around a sample's mean value.

Example 5.10

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

SOLUTION

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

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

and that the square of the residual error is

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

The following table displays the results for all six solutions.

x_i	y_i	$\widehat{y}_i$	$(y_i - \widehat{y}_i)^2$
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

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

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

$$s_{b_1} = \sqrt{\frac{n s_r^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = \sqrt{\frac{6 \times (0.4035)^2}{(6 \times 0.550) - (1.500)^2}} = 0.965$$

$$s_{b_0} = \sqrt{\frac{s_r^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = \sqrt{\frac{(0.4035)^2 \times 0.550}{(6 \times 0.550) - (1.500)^2}} = 0.292$$

You can find values for t in Appendix 4.

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

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

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

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, we evenly space our standards over a wide range of analyte concentrations. A close examination of equation 5.20 and equation 5.21 help us appreciate why this is true. The denominators of both equations include the term $\sum (x_i - \bar{x})^2$. The larger the value of this term—which we 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 helps to decrease the value of the term $\sum x_i$ in equation 5.21, which we 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 we use 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

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

where m is the number of replicate we use 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 the 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 is based on the desired level of confidence and n–2 degrees of freedom.

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.

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

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.

Example 5.11

Three replicate analyses for a sample that contains an unknown concentration of analyte, yield values for S_{samp} of 29.32, 29.16 and 29.51 (arbitrary units). 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, $\overline{S}_{samp}$, 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

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

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

$$\sum_{i=1}^{n} (C_{std_i} - \overline{C}_{std})^2 = (s_{C_{std}})^2 \times (n - 1)$$

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.

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}.

You can find values for t in Appendix 4.
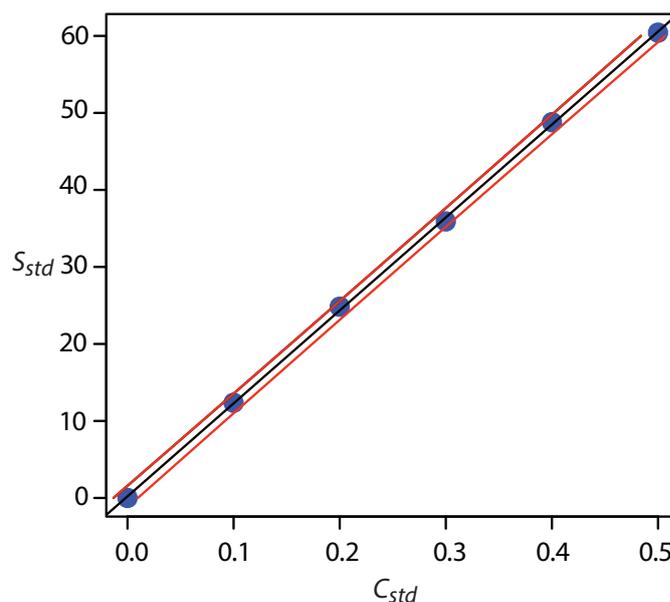
where $s_{C_{std}}$ is the standard deviation for the concentration of analyte in the calibration standards. Using the data in Table 5.1 we find that $s_{C_{std}}$ is 0.1871 and

$$\sum_{i=1}^{n}(C_{std_i} - \overline{C}_{std})^2 = (0.1872)^2 \times (6-1) = 0.175$$

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

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

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

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

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

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

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.

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 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

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

If the regression model is valid, then the residual errors should be distributed randomly 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 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, 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 errors are not random, which suggests we cannot model the data using a straight-line relationship. Regression methods for the latter two cases are discussed in the following sections.
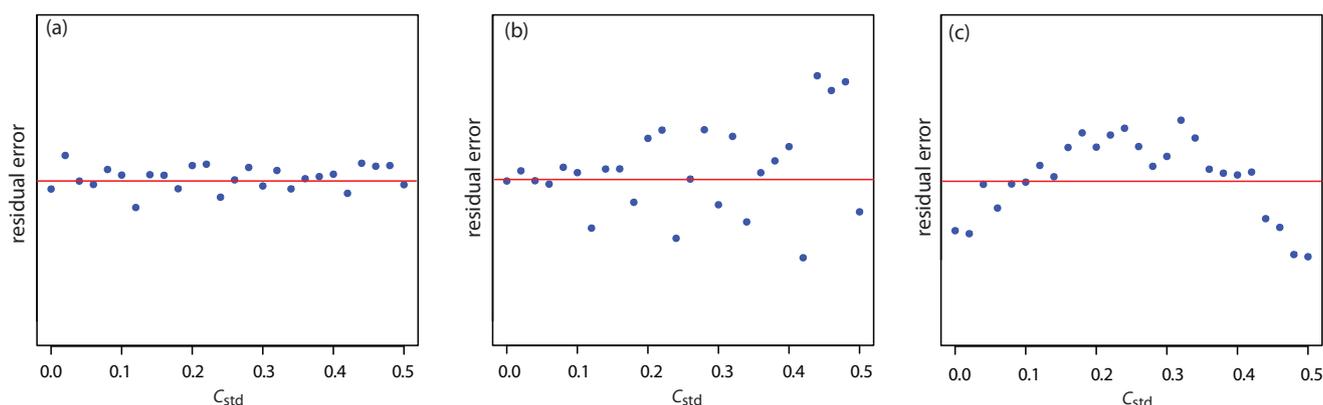
Figure 5.13 Plots of the **residual error** in the signal, S_{std}, as a function of the concentration of analyte, C_{std}, for an unweighted straight-line regression model. The **red** line shows a residual error of zero. The distribution of the residual errors in (a) indicates that the unweighted linear regression model is appropriate. The increase in the residual errors in (b) for higher concentrations of analyte, suggests 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.

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_0, and the slope, b_1; thus

$$b_0 = \frac{\sum_{i=1}^{n} w_i y_i - b_1 \sum_{i=1}^{n} w_i x_i}{n} \qquad 5.26$$

$$b_1 = \frac{n \sum_{i=1}^{n} w_i x_i y_i - \sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i y_i}{n \sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{i=1}^{n} w_i x_i\right)^2} \qquad 5.27$$

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

$$w_i = \frac{n(s_{y_i})^{-2}}{\sum_{i=1}^{n} (s_{y_i})^{-2}} \qquad 5.28$$

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.

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.

SOLUTION

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

x_i	y_i	s_{y_i}	$(s_{y_i})^{-2}$	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

Adding together the values in the forth column gives

$$\sum_{i=1}^{n}(s_{y_i})^{-2}$$

which we use to calculate the individual weights in the last column. After we calculate 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

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

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

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.

$$\sum_{i=1}^{n} w_i x_i = 0.3644 \qquad \sum_{i=1}^{n} w_i y_i = 44.9499$$

$$\sum_{i=1}^{n} w_i x_i^2 = 0.0499 \qquad \sum_{i=1}^{n} w_i x_i y_i = 6.1451$$

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.
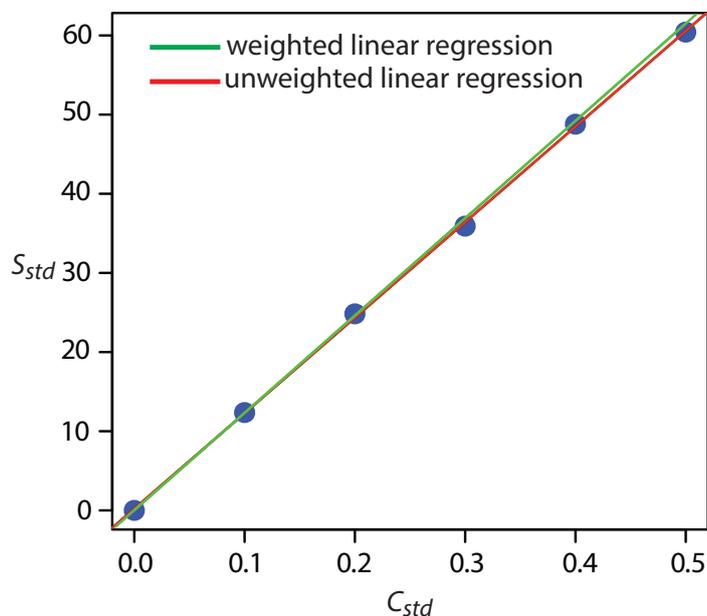

Figure 5.14 A comparison of the **unweighted** and the **weighted** normal calibration curves. See Example 5.9 for details of the unweighted linear regression and Example 5.12 for details of the weighted linear regression.

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

$$y_c = \frac{1}{n}\sum_{i=1}^{n} w_i x_i$$

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

If we remove our assumption that indeterminate errors affecting a calibration curve are present only in the signal (y), then we also must factor into the regression model the indeterminate errors that affect 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.

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, of which the most obvious is that data with a uniform variance in y will not maintain that uniform variance after it is transformed.

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.

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

It is worth noting that the term "linear" does not mean a straight-line. A linear function may contain more than one additive term, but each such term has one and only one adjustable multiplicative parameter. The function

$$y = ax + bx^2$$

is an example of a linear function because the terms x and x^2 each include a single multiplicative parameter, a and b, respectively. The function

$$y = x^b$$

is nonlinear because b is not a multiplicative parameter; it is, instead, a power. 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 into a linear function. For example, taking the log of both sides of the nonlinear function above gives a linear function.

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

8 Bonate, P. J. *Anal. Chem.* **1993**, *65*, 1367–1372.
9 See, for example, Analytical Methods Committee, "Fitting a linear functional relationship to data with error on both variable," AMC Technical Brief, March, 2002), as well as this chapter's Additional Resources.
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.

The regression models in this chapter apply only to functions that contain 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

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

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

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

5E   Compensating for the Reagent Blank (S_{reag})

Thus far in our discussion of strategies for standardizing analytical methods, we have assumed that a suitable reagent blank is available 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 with different sizes, 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.

---

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

In working up this data, the analytical chemists used at least four different approaches to correct the 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. The first four rows of Table 5.4 shows the equations for calculating the analyte's concentration using each approach, along with the reported concentrations 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 and any bias that results 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 that results 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, it must be determined using the sample itself. One approach is to measure the signal for samples of differ-

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.

ent size, and to determine the regression line for a plot of S_{samp} versus the 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 is 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.

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

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:

$$= \text{intercept}(known\_y's, known\_x's)$$

$$= \text{slope}(known\_y's, known\_x's)$$

	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}).

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

where *known_y's* is the range of cells that contain the signals (y), and *known_x's* is the range of cells that contain the concentrations (x). For example, if you click on an empty cell and enter

$$= \text{slope(B2:B7, A2:A7)}$$

Excel returns 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 Data 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.

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.

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 be familiar. The correlation coefficient is a measure of the extent to which the regression model explains the variation in y. Values of r

Excel's Data Analysis Toolpak is available for Windows. Older versions of Excel for Mac included the toolpak; however, beginning with Excel for Mac 2011, the toolpak no longer is available.

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

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

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.
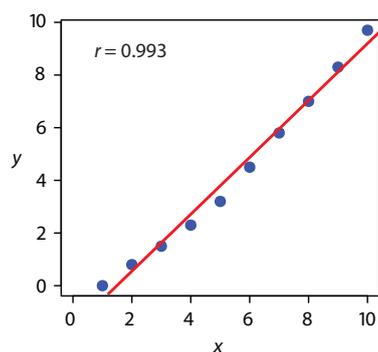
Figure 5.17 Example of fitting a straight-line (in **red**) to curvilinear data (in **blue**).

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 there is no relationship between x and y. In developing the calculations for linear regression, we did not consider the correlation coefficient. There is a reason for this. For most straight-line calibration curves the correlation coefficient is 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 useful counterexample. Although the regression line has a correlation coefficient of 0.993, the data clearly is curvilinear. The take-home lesson here is simple: do not 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.

H_0: the regression model does not explain the variation in y

H_A: the 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 2.5×10^{-6}%, which 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 9.0×10^{-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 C_{std}. 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.

slope H_0: $\beta_1 = 0$, H_A: $\beta_1 \neq 0$

y-intercept H_0: $\beta_0 = 0$, H_A: $\beta_0 \neq 0$

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

	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 shaded cells 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.

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

$$=\text{sum}(\textit{first cell}:\textit{last cell})$$

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

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 **Charts** from the ribbon. Select **Scatter**, choosing the option without lines that connect the points. To add a regression line to the chart, click on the chart's data 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 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 to calculate the uncertainty when predicting values of *x*. In terms of this chapter, Excel can not calculate the uncertainty for the analyte's
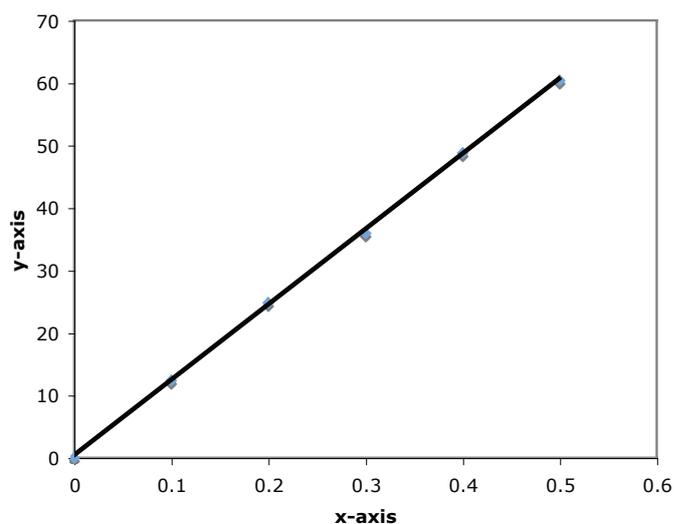
Figure 5.19 Example of an Excel scatterplot showing the data and a regression line.

concentration, C_A, given the signal for a sample, S_{samp}. Another limitation is that Excel does not have a built-in function for a weighted linear regression. You can, however, program a spreadsheet to handle these calculations.

5F.2   R

Let's use R 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 that contain 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)
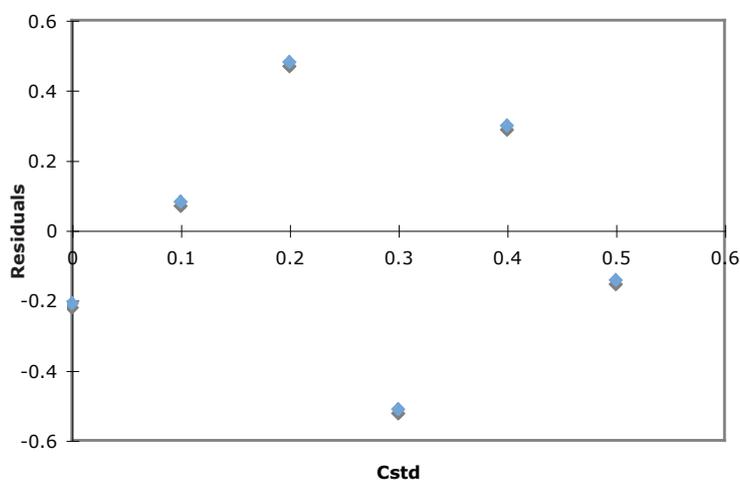
Figure 5.20 Example of Excel's plot of a regression model's residual errors.

The command for a straight-line linear regression model is

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

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

> model $=$ lm(signal $\sim$ conc)

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

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 the objects that contain our data, and the **abline** command

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

where *object* is the object that contains 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.

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

> summary(model)

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

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

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

$$y = a + bx$$

where a is the y-intercept and b is the slope.
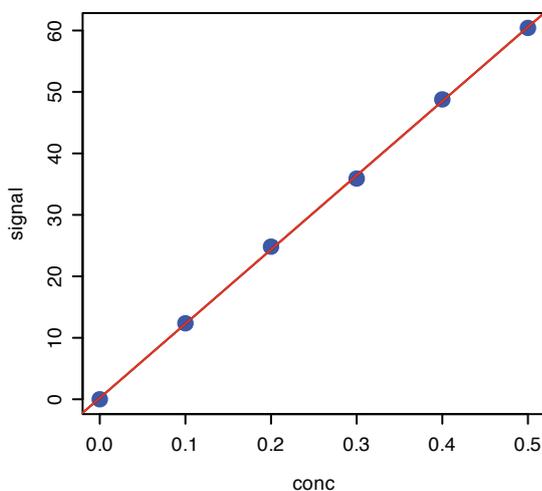
Figure 5.21 Example of a regression plot in R showing the data (in **blue**)and the regression line (in **red**). You can customize your plot by adjusting the plot command's optional arguments. For example, the argument *pch* controls the symbol used for plotting points, the argument *col* allows you 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

$$\text{help(plot)}$$

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.

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.

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)

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.
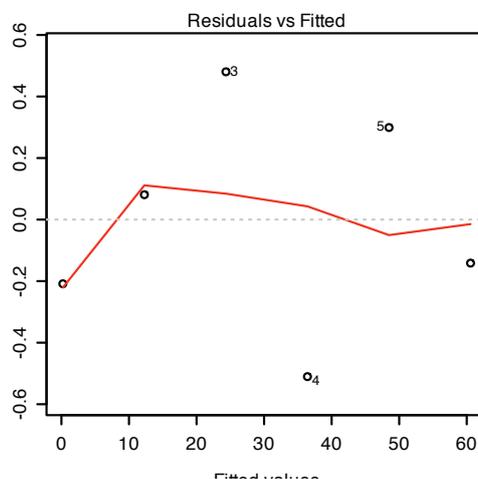

Figure 5.23 Example showing R's plot of a regression model's residual error.

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

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

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

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.

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 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*).

> 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 that contains the regression model's results, *newdata* is an object that contains 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 that contains 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, $\pm t s_{C_A}$, for the analyte's concentration, and *$`Confidence Limits`* provides the lower limit and upper limit for the confidence interval for C_A.

```
> 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}.

USING R FOR A WEIGHTED LINEAR REGRESSION

R's command for an unweighted linear regression also allows for a weighted linear regression if we include an additional argument, *weights*, whose value is an object that contains the weights.

$$lm(y \sim x, weights = object)$$

Let's use this command to complete Example 5.12. First, we need to create an object that contains 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 that contains 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.

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.

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

calibration curve	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

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 that affect 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 use 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 available to us, including external standards, the method of standard addition, and internal standards. The

most common strategy is a multiple-point external standardization and a normal calibration curve. We use the method of standard additions, in which we add known amounts of analyte 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 provides 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, are used.

5I   Problems

1. Suppose you 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 prepare each solution as 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 ± 0.0002.

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

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 the 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 the internal standard of 0.274 and 0.198, respectively. Report the analyte's concentration in the sample.

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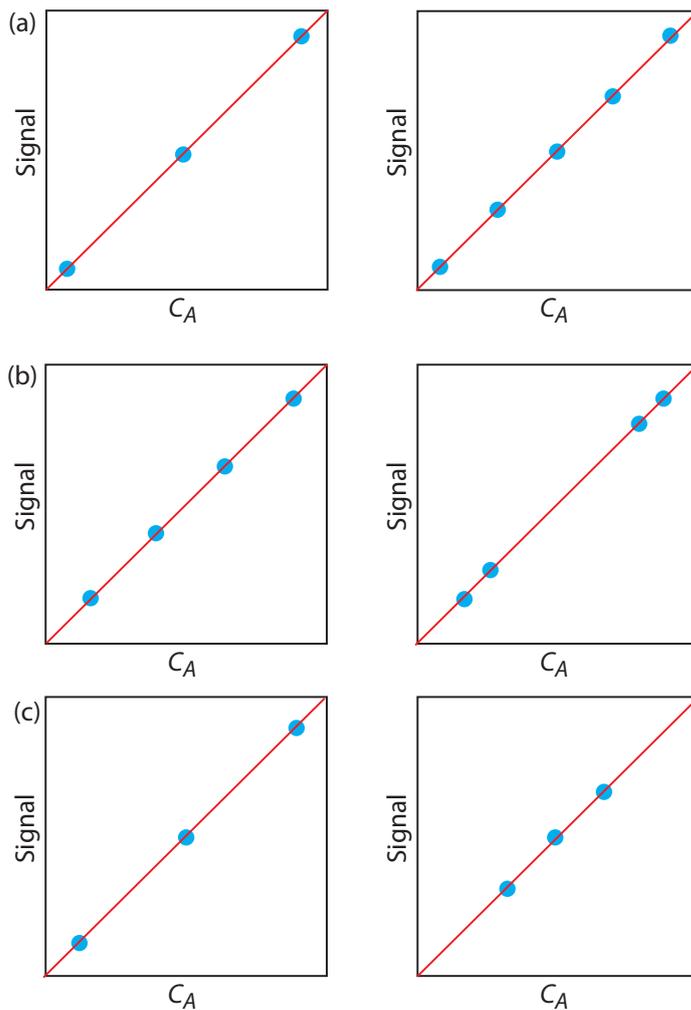

Figure 5.26 Calibration curves to accompany Problem 7.

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_{spike} (nA)	4.8	11.4	18.2	26.6	32.3	37.7

(a) Use a linear regression analysis to determine the equation for the calibration curve 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_{spike} (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 nA. 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 addition is performed. A 5.00-mL portion of sample is analyzed and then successive 0.10-mL spikes of a 600.0 ppb standard of the analyte are 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	2.080	2.550

14 Wojciechowski, M.; Balcerzak, J. *Anal. Chim. Acta* **1991**, *249*, 433–445.
15 Troost, J. R.; Olavesen, E. Y. *Anal. Chem.* **1996**, *68*, 708–711.

(a) Determine the equation for the calibration curve using a linear regression, and report confidence intervals for the slope and the y-intercept. Average the replicate signals for each standard before you complete the linear regression analysis.

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

11. In Chapter 4 we used a paired t-test to compare two analytical methods that were used to analyze independently 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.

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

Rearranging equation 5.22 gives the test statistic for the slope.

$$t_{exp} = \frac{|\beta_1 - b_1|}{s_{b_1}} = \frac{|1 - b_1|}{s_{b_1}}$$

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

12. Consider the following three data sets, each of which gives values 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

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 an 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.

These three data sets are taken from Anscombe, F. J. "Graphs in Statistical Analysis," *Amer. Statis.* **1973**, *27*, 17-21.

(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. 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 the 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.

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

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

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

When using multiple standards, the indeterminate errors that affect the signal for one standard are partially compensated for by the indeterminate errors that affect 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

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

If we plug this result into the equation for the x-intercept and solve for C_A, we find that the concentration of Mn^{2+} is

$$C_A = -\frac{(x\text{-intercept})C_{std}}{V_o} = -\frac{(-1.731\,\text{mL}) \times 100.6\,\text{mg/L}}{25.00\,\text{mL}} = 6.96\,\text{mg/L}$$

For Figure 7b, the x-intercept is

$$x\text{-}int ercept = -\frac{0.1478}{0.0425 \text{ mL/mg}} = -3.478 \text{ mg/mL}$$

and the concentration of Mn^{2+} is

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

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	$x_i y_i$	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

$$\sum_{i=1}^{n} x_i = 2.371 \times 10^{-2} \quad \sum_{i=1}^{n} y_i = 0.710$$

$$\sum_{i=1}^{n} x_i y_i = 4.110 \times 10^{-3} \quad \sum_{i=1}^{n} x_i^2 = 1.378 \times 10^{-4}$$

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

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

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

and that the regression equation is

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

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

x_i	y_i	$\widehat{y}_i$	$(y_i - \widehat{y}_i)^2$
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

$$s_r = \sqrt{\frac{1.596 \times 10^{-5}}{6 - 2}} = 1.997 \times 10^{-3}$$

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

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

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

and use them to calculate the 95% confidence intervals for the slope and the y-intercept

$$\beta_1 = b_1 \pm ts_{b_1} = 29.57 \pm (2.78 \times 0.3007) = 29.57 \text{ M}^{-1} \pm 0.84 \text{ M}^{-1}$$

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

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

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

The standard deviation in C_A is

$$s_{C_A} = \frac{1.997 \times 10^{-3}}{29.57} \sqrt{\frac{1}{3} + \frac{1}{6} + \frac{(0.114 - 0.1183)^2}{(29.57)^2 \times (4.408 \times 10^{-5})}}$$
$$= 4.778 \times 10^{-5}$$

and the 95% confidence interval is

$$\mu = C_A \pm ts_{C_A} = 3.80 \times 10^{-3} \pm \{2.78 \times (4.778 \times 10^{-5})\}$$

$$\mu = 3.880 \times 10^{-3} \text{ M} \pm 0.13 \times 10^{-3} \text{ M}$$

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	$\widehat{y}_i$	$y_i - \widehat{y}_i$
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
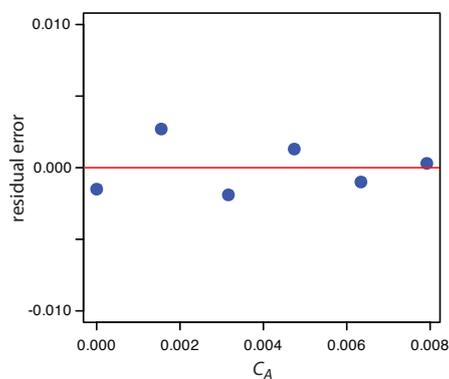
Figure 5.27 Plot of the residual errors for the data in Practice Exercise 5.5.

Figure 5.27 shows a plot of the resulting residual errors. The residual errors appear random, although they do alternate in sign, and that 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, provides 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, we find that C_A is

$$C_A = \frac{S_{samp} - b_0}{b_1} = \frac{0.114 - 0.0014}{29.59 \ \mathrm{M}^{-1}} = 3.80 \times 10^{-3} \ \mathrm{M}$$

The standard deviation in C_A is

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.99979366
R Square	0.99958737
Adjusted R Sq	0.99948421
Standard Erro	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.

$$s_{C_A} = \frac{1.996 \times 10^{-3}}{29.59} \sqrt{\frac{1}{3} + \frac{1}{6} + \frac{(0.114 - 0.1183)^2}{(29.59)^2 \times (4.408 \times 10^{-5})}}$$

$$= 4.772 \times 10^{-5}$$

and the 95% confidence interval is

$$\mu = C_A \pm t s_{C_A} = 3.80 \times 10^{-3} \pm \{2.78 \times (4.772 \times 10^{-5})\}$$

$$\mu = 3.80 \times 10^{-3}\ M \pm 0.13 \times 10^{-3}\ M$$

Click <u>here</u> to return to the chapter

Practice Exercise 5.7

Figure 5.29 shows the R session for this problem, including loading the *chemCal* package, creating objects to hold the values for C_{std}, S_{std}, and S_{samp}. Note that for S_{samp}, 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 <u>here</u> 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 <u>Practice Exercise 5.7</u>.