

# Chapter 4

## Evaluating Analytical Data

### Chapter Overview

- 4A Characterizing Measurements and Results
- 4B Characterizing Experimental Errors
- 4C Propagation of Uncertainty
- 4D The Distribution of Measurements and Results
- 4E Statistical Analysis of Data
- 4F Statistical Methods for Normal Distributions
- 4G Detection Limits
- 4H Using Excel and R to Analyze Data
- 4I Key Terms
- 4J Chapter Summary
- 4K Problems
- 4L Solutions to Practice Exercises

When using an analytical method we make three separate evaluations of experimental error. First, before beginning an analysis we evaluate potential sources of errors to ensure that 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, comparing 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](http://www.usmint.gov)).

## 4A Characterizing Measurements and Results

Let’s begin by choosing a simple quantitative problem requiring a single measurement—What is the mass of a penny? As you consider this question, you probably recognize that it is too broad. 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 limit our problem to pennies from the United States.

There are other concerns we might consider. For example, the United States Mint currently produces pennies at two locations (Figure 4.1). Because it seems unlikely that a penny’s mass depends upon 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 narrow our question to—What is the mass of a circulating United States Penny?

A good way to begin our analysis is to examine some preliminary data. Table 4.1 shows masses for seven pennies from my change jar. In examining this data it is immediately apparent 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 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 are randomly scattered around a central value that provides 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**,  $\bar{X}$ , 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

$$\bar{X} = \frac{\sum X_i}{n}$$

where  $X_i$  is the  $i^{\text{th}}$  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

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

The mean is the most common estimator of central tendency. It is not a robust estimator, however, because an extreme value—one much larger or much smaller than the remainder of the data—strongly influences the mean's value.<sup>1</sup> For example, if we mistakenly 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 estimator is robust if its value is not affected too much by an unusually large or unusually small measurement.

#### **MEDIAN**

The **MEDIAN**,  $\tilde{X}$ , is the middle value when we order our data from the smallest to the largest value. When the data set includes an odd number of entries, the median is the middle value. For an even number of entries, 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

$$3.056 \quad 3.080 \quad 3.094 \quad 3.107 \quad 3.112 \quad 3.174 \quad 3.198$$

Because there are seven measurements, the median is the fourth value in the ordered data set; 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 compara-

<sup>1</sup> Rousseeuw, P. J. *J. Chemom.* **1991**, *5*, 1–20.

ble in magnitude. The median, however, provides a more robust estimate of central tendency because it is less sensitive to measurements with extreme values. For example, introducing the transcription error discussed earlier for the mean changes the median's value from 3.107 g to 3.112 g.

## 4A.2 Measures of Spread

If the mean or median provides an estimate of a penny's expected mass, then the spread of individual measurements provides an estimate of the difference in mass among pennies or of the uncertainty in measuring mass with a balance. Although we often define spread relative to a specific measure of central tendency, its magnitude is independent of the central value. Changing all measurements in the same direction, by adding or subtracting a constant value, changes the mean or median, but 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 any 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 a data set's individual values about its mean, and is given as

$$s = \sqrt{\frac{\sum_i (X_i - \bar{X})^2}{n - 1}} \quad 4.1$$

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

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

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

### Example 4.3

What are the standard deviation, the relative standard deviation and the percent relative standard deviation for the data in [Table 4.1](#)?

[Problem 12](#) at the end of the chapter asks you to show that this is true.

**SOLUTION**

To calculate the standard deviation we first calculate the difference between each measurement and the mean value (3.117), square the resulting differences, and add them together to give the numerator of [equation 4.1](#).

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

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

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

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

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

$$s_r = \frac{0.051 \text{ g}}{3.117 \text{ g}} = 0.016 \quad \%s_r = (0.016) \times 100\% = 1.6\%$$

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

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 square of the standard deviation, or the **VARIANCE**. We usually report a data set's standard deviation, rather than its variance, because the mean value and the standard deviation have 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<sup>+</sup> 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 [here](#) to review your answer to this exercise.

## 4B Characterizing Experimental Errors

Characterizing the mass of a penny using the data in [Table 4.1](#) 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 asks about their precision. In this section we consider the types of experimental errors affecting accuracy and precision.

### 4B.1 Errors Affecting Accuracy

Accuracy is a measure of how close a measure of central tendency is to the expected value,  $\mu$ . We can express accuracy as either an absolute error,  $e$

$$e = \bar{X} - \mu \quad 4.2$$

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

$$\%e_r = \frac{\bar{X} - \mu}{\mu} \times 100 \quad 4.3$$

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

We call errors affecting the accuracy of an analysis determinate. Although there may be several different sources of **DETERMINATE ERROR**, each source 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. The cumulative effect of these determinate errors is a net positive or negative 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 representative sample. For example, if you monitor the envi-

The convention for representing statistical parameters is to use a Roman letter for a value calculated from experimental data, and a Greek letter for the corresponding expected value. For example, the experimentally determined mean is  $\bar{X}$ , and its underlying expected value is  $\mu$ . Likewise, the standard deviation by experiment is  $s$ , and the underlying expected value is  $\sigma$ .

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.

ronmental quality of a lake by sampling a single location near a point source of pollution, such as an outlet for industrial effluent, then your results will be misleading. In determining the mass of a U. S. penny, our strategy for selecting pennies must ensure that we do not include pennies from other countries.

## METHOD ERRORS

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

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

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

where  $k_A$  is the method's sensitivity for the analyte and  $S_{\text{mb}}$  is the signal from the method blank. A determinate **METHOD ERROR** exists when our value for  $k_A$  or  $S_{\text{mb}}$  is invalid. For example, a method in which  $S_{\text{total}}$  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  $\pm 0.02$  mL, which means that 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; thus, the pipet's expected volume is a fixed value within the stated range.

Volumetric glassware is categorized into classes depending on its accuracy. Class A glassware is manufactured to comply with tolerances specified by agencies 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 we normally can use it without calibration. The tolerance levels for Class B glassware are usually twice those for Class A glassware. Other types of volumetric glassware, such as beakers and graduated cylinders, are unsuitable for accurately measuring volumes. [Table 4.2](#) provides a summary of typical measurement errors for Class A volumetric glassware. Tolerances for digital pipets and for balances are listed in [Table 4.3](#) and [Table 4.4](#).

We can minimize determinate measurement errors by calibrating our equipment. Balances are calibrated using a reference weight whose mass can

An awareness of potential sampling errors is especially important when working with heterogeneous materials. Strategies for obtaining representative samples are covered in [Chapter 5](#).



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

be traced back to the SI standard kilogram. Volumetric glassware and digital pipets can be calibrated by determining the mass of water that it delivers or contains and using the density of water to calculate the actual volume. It is never safe to assume that a calibration will remain unchanged 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, leads to small, but significant changes in the glassware's calibration.<sup>2</sup> Many instruments drift out of calibration over time and may require frequent recalibration during an analysis.

<sup>2</sup> 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<sup>†</sup>**

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		

<sup>†</sup> Tolerance values are from the ASTM E288, E542, and E694 standards.

**Table 4.3 Measurement Errors for Digital Pipets<sup>†</sup>**

Pipet Range	Volume (mL or $\mu\text{L}$ ) <sup>‡</sup>	Percent Measurement Error
10–100 $\mu\text{L}$	10	±3.0%
	50	±1.0%
	100	±0.8%
100–1000 $\mu\text{L}$	100	±3.0%
	500	±1.0%
	1000	±0.6%
1–10 mL	1	±3.0%
	5	±0.8%
	10	±0.6%

<sup>†</sup> Values are from [www.eppendorf.com](http://www.eppendorf.com). <sup>‡</sup> 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	$\pm 1$ mg
A & D ER 120M	120	$\pm 0.1$ mg
Metler H54	160	$\pm 0.01$ mg

### PERSONAL ERRORS

Finally, analytical work is always subject to **PERSONAL ERROR**, including the ability to see a change in the color of an indicator signaling 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.

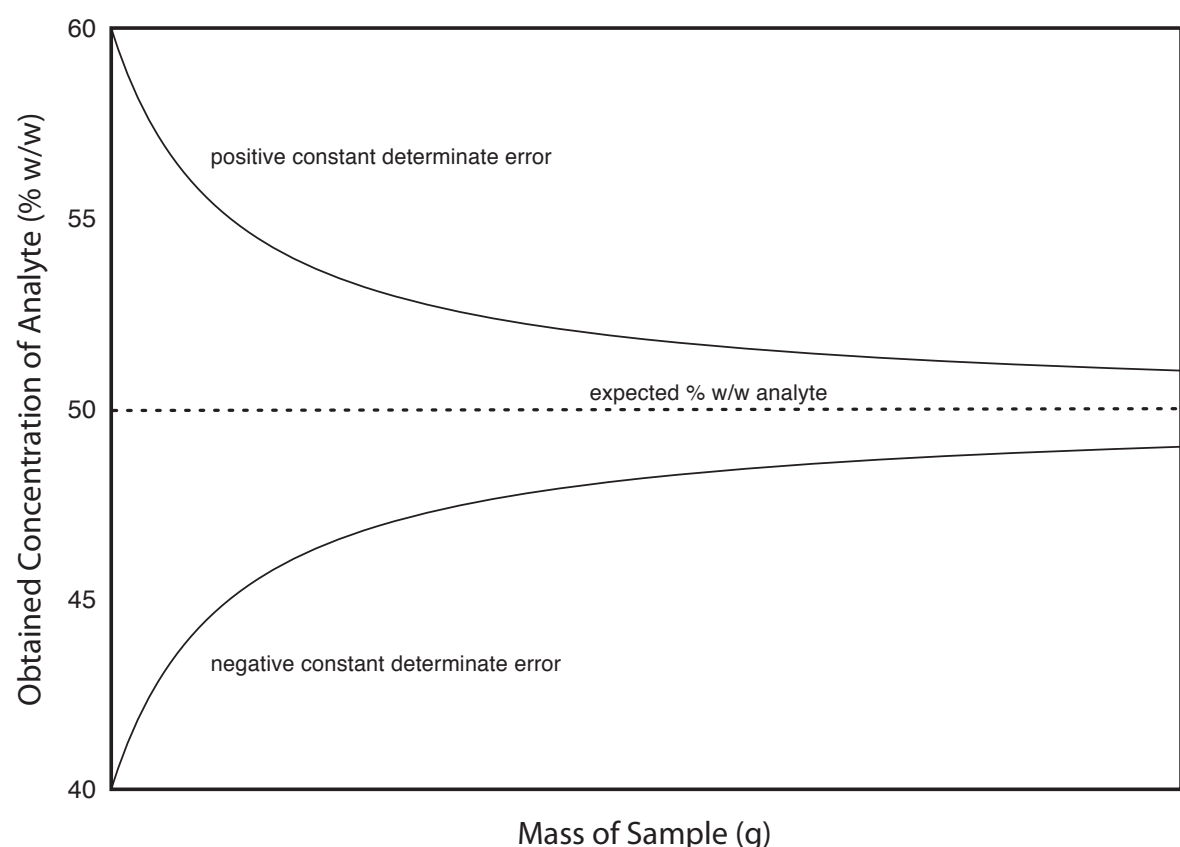
### IDENTIFYING DETERMINATE ERRORS

Determinate errors can be difficult to detect. Without knowing the expected value for an analysis, the usual situation in any analysis that matters, there is 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 analyzing 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 that the sample is 50.0% w/w analyte. As shown 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 obtained results become closer to the expected result. An upward or downward trend in a graph of the analyte's obtained

Table 4.5 Effect of a Constant Determinate Error on the Analysis of a Sample Containing 50% w/w Analyte

Mass Sample (g)	Expected Mass of Analyte (g)	Constant Error (g)	Obtained Mass of Analyte (g)	Obtained 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 determinate error on the determination of an analyte in samples of varying size.

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 showing 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 containing a known amount of analyte in a matrix similar to the 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 *Ginkgo bilboa* leaves. Another approach is to compare your analysis to an analysis carried out using an independent analytical method 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 Containing 50% w/w Analyte

Mass Sample (g)	Expected Mass of Analyte (g)	Proportional Error (%)	Obtained Mass of Analyte (g)	Obtained 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)<sup>†</sup>

Class of Analyte	Analyte	Mass Fraction (mg/g or ng/g)		
<b>Flavonoids/Ginkgolide B</b> (mass fractions in mg/g)	Quercetin	2.69	±	0.31
	Kaempferol	3.02	±	0.41
	Isorhamnetin	0.517	±	0.099
	Total Aglycones	6.22	±	0.77
<b>Selected Terpenes</b> (mass fractions in mg/g)	Ginkgolide A	0.57	±	0.28
	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
<b>Selected Toxic Elements</b> (mass fractions in ng/g)	Cadmium	20.8	±	1.0
	Lead	995	±	30
	Mercury	23.08	±	0.17

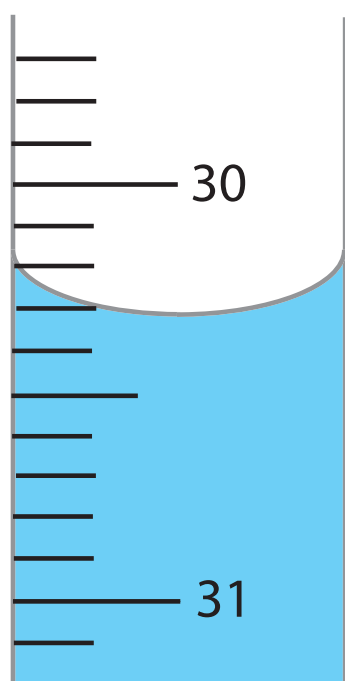
<sup>†</sup> The primary purpose of this Standard Reference Material is to validate analytical methods for determining flavonoids, terpene lactones, and toxic elements in *Ginkgo biloba* or other materials with a similar matrix. Values are from the official Certificate of Analysis available at [www.nist.gov](http://www.nist.gov).

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 a constant value to the signal. A poorly calibrated method, which yields an invalid sensitivity for the analyte,  $k_A$ , will result in a proportional determinate error.

## 4B.2 Errors Affecting Precision

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. We make a distinction between two types of precision: repeatability and reproducibility. **REPEATABILITY** is the precision when a single analyst completes the 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 affecting 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,



**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  $\pm 0.01$ – $0.03$  mL.

provided that enough measurements are made. In such situations the mean or median is largely unaffected by the precision of the analysis.

### SOURCES OF INDETERMINATE ERROR

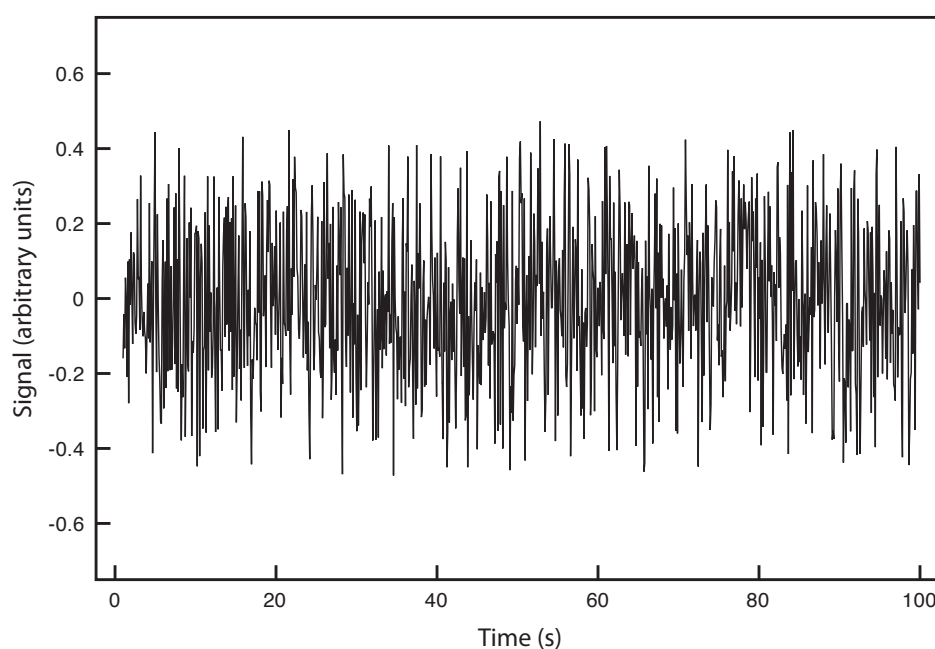
We can assign indeterminate errors to several sources, including collecting samples, manipulating samples during the analysis, and making measurements. When collecting a sample, for instance, only a small portion of the available material is taken, increasing the chance that small-scale inhomogeneities in the sample will affect repeatability. Individual pennies, for example, may show variations 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 for introducing 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 introduces an indeterminate method error.

Finally, any measuring device is subject to an indeterminate measurement error due to limitations in reading its scale. For example, a buret with scale divisions every 0.1 mL has an inherent indeterminate error of  $\pm 0.01$ – $0.03$  mL when we estimate the volume to the hundredth of a milliliter (Figure 4.4).

### EVALUATING INDETERMINATE ERROR

An indeterminate error due to analytical equipment or instrumentation is generally easy to estimate by measuring the standard deviation for several replicate measurements, or by monitoring the signal's fluctuations over time in the absence of analyte (Figure 4.5) and calculating 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 indeterminate measurement error on our analysis of the mass of a circulating United States penny, we might make several determinations for the mass of 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 determining 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 significant 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.<sup>3</sup> **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 can divide error into determinate and indeterminate sources. Although we can correct for determinate errors, the indeterminate portion of the error remains. With statistical significance testing, which is discussed later in this chapter, we can determine if our results show evidence of bias.

**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, providing an estimate of indeterminate errors. Uncertainty accounts for all errors—both determinate and indeterminate—that might reasonably affect a measurement or result. Although we always try to correct determinate errors before beginning 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  $\pm 0.02$  mL is its uncertainty because your best estimate of its expected volume is  $10.00 \text{ mL} \pm 0.02 \text{ mL}$ . This uncertainty is primarily determinate error. If you use the pipet to dispense several replicate portions of solution, 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  $\pm 0.006$  mL. This standard deviation is the precision

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

<sup>3</sup> 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

with which we expect to deliver a solution using a Class A 10-mL pipet. In this case the published uncertainty for the pipet ( $\pm 0.02$  mL) is worse than its experimentally determined precision ( $\pm 0.006$  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 this pipet's expected volume, then its uncertainty is  $\pm 0.006$  mL. As expected, calibrating the pipet allows us to decrease its uncertainty.<sup>4</sup>

## 4C Propagation of Uncertainty

Suppose you 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  $9.992 \pm 0.006$  mL, what is the volume and uncertainty when 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 previous discussion we know that the total uncertainty is greater than  $\pm 0.000$  mL and less than  $\pm 0.012$  mL. To estimate the cumulative effect of multiple uncertainties 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.

<sup>4</sup> Kadis, R. *Talanta* **2004**, *64*, 167–173.

Although we will not derive or further justify these rules here, you may consult the additional resources at the end of this chapter for references that discuss the propagation of uncertainty in more detail.

## 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 the result. For the equations in this section we represent the result with the symbol  $R$ , and 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 adding or subtracting measurements we use their absolute uncertainties for a propagation of uncertainty. For example, if the result is given by the equation

$$R = A + B - C$$

then the absolute uncertainty in  $R$  is

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

### Example 4.5

When dispensing 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 mL  $\pm$  0.02 mL, and then using the calibration data from [Table 4.9](#).

#### **SOLUTION**

To calculate the total volume we simply 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 value 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 mL  $\pm$  0.03 mL when using the tolerance values, and 19.98  $\pm$  0.01 mL when using the calibration data.

### 4C.3 Uncertainty When Multiplying or Dividing

When multiplying or dividing measurements we use their relative uncertainties for a propagation of uncertainty. For example, if the result is given by the equation

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

then the relative uncertainty in R is

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

#### Example 4.6

The quantity of charge,  $Q$ , in coulombs passing 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 A  $\pm$  0.01 A passes through the circuit for 120 s  $\pm$  1 s, what is the total charge passing through the circuit 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

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

The absolute uncertainty in the charge 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 C  $\pm$  1 C.



## 4C.4 Uncertainty for Mixed Operations

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

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

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

#### SOLUTION

Rearranging the equation and solving for  $C_A$

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

gives the analyte's concentration as 126 ppm. To estimate the uncertainty in  $C_A$ , we first determine the uncertainty for the numerator using [equation 4.6](#).

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

The numerator, therefore, is  $23.41 \pm 0.028$ . To complete the calculation we estimate the relative uncertainty in  $C_A$  using [equation 4.7](#).

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

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  $126 \text{ ppm} \pm 2 \text{ ppm}$ .

### Practice Exercise 4.2

To prepare a standard solution of  $\text{Cu}^{2+}$  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  $\text{HNO}_3$ , 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  $\text{Cu}^{2+}$  in mg/L, and its uncertainty? Assume that the uncertainty in the balance is  $\pm 0.1 \text{ mg}$  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<sup>†</sup>**

Function	$u_R$
$R = kA$	$u_R = ku_A$
$R = A + B$	$u_R = \sqrt{u_A^2 + u_B^2}$
$R = A - B$	$u_R = \sqrt{u_A^2 + u_B^2}$
$R = A \times B$	$\frac{u_R}{R} = \sqrt{\left(\frac{u_A}{A}\right)^2 + \left(\frac{u_B}{B}\right)^2}$
$R = \frac{A}{B}$	$\frac{u_R}{R} = \sqrt{\left(\frac{u_A}{A}\right)^2 + \left(\frac{u_B}{B}\right)^2}$
$R = \ln(A)$	$u_R = \frac{u_A}{A}$
$R = \log(A)$	$u_R = 0.4343 \times \frac{u_A}{A}$
$R = e^A$	$\frac{u_R}{R} = u_A$
$R = 10^A$	$\frac{u_R}{R} = 2.303 \times u_A$
$R = A^k$	$\frac{u_R}{R} = k \times \frac{u_A}{A}$

<sup>†</sup> Assumes that the measurements  $A$  and  $B$  are independent;  $k$  is a constant whose value has no uncertainty.

### 4C.5 Uncertainty for Other Mathematical Functions

Many other mathematical operations are common in analytical chemistry, including 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  $\pm 0.03$ , what is the  $[H^+]$  and its uncertainty?



**SOLUTION**

The concentration of  $\text{H}^+$  is

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

or  $1.9 \times 10^{-4} \text{ M}$  to two significant figures. From [Table 4.10](#) the relative uncertainty in  $[\text{H}^+]$  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  $[\text{H}^+]$  as  $1.9 (\pm 0.1) \times 10^{-4} \text{ M}$ .

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

$$A = -\log \frac{P}{P_0}$$

where  $P_0$  is the power of radiation from the light source and  $P$  is the power after it passes through the solution. What is the absorbance if  $P_0$  is  $3.80 \times 10^2$  and  $P$  is  $1.50 \times 10^2$ ? If the uncertainty in measuring  $P_0$  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 for completing 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 mass twice—once to tare the balance at 0.000 g, and once to measure the penny's mass. If the uncertainty for measuring mass is  $\pm 0.001 \text{ g}$ , then we estimate the uncertainty in measuring mass as

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

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

We also can use 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  $126 \text{ ppm} \pm 2 \text{ ppm}$ , 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)

$$\frac{2 \text{ ppm}}{126 \text{ ppm}} \times 100 = 1.6\%$$

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

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 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  $\pm 0.0015 \text{ ppm}^{-1}$ .

#### Practice Exercise 4.4

Verify that an uncertainty of  $\pm 0.0015 \text{ ppm}^{-1}$  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 diluting a stock solution there are usually several different 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 by 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 transfer pipets and volumetric flasks. As shown below, we can use the tolerance values for volumetric glassware to determine the optimum dilution strategy.<sup>5</sup>

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

### 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 using a 1-mL pipet and a 1000-mL volumetric flask.
- (b) A two-step dilution using 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.

See [Appendix 2](#) for a more detailed treatment of the propagation of uncertainty.

## 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 experiments raises some interesting questions. Are the results for one experiment better than those 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 to predict the properties of all pennies by analyz-

**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		
$\bar{X}$	3.117		3.081
$s$	0.051		0.037

ing 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 our experiment, 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, are results for two samples drawn from the larger population of all circulating United States pennies.

#### 4D.2 Probability Distributions for Populations

Table 4.11 provides the mean and standard deviation 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 something about how the masses of individual pennies are distributed around the population's average mass. We represent the distribution of a population by plotting the probability or frequency of obtaining an 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

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

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 that 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 can take on 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 obeying 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  $N \times (N-1) \times (N-2) \times \dots \times 1$ . 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,  $^{12}\text{C}$  and  $^{13}\text{C}$ , with relative isotopic abundances of, respectively, 98.89% and 1.11%.

- What are the mean and the standard deviation for the number of  $^{13}\text{C}$  atoms in a molecule of cholesterol ( $\text{C}_{27}\text{H}_{44}\text{O}$ )?
- What is the probability that a molecule of cholesterol has no atoms of  $^{13}\text{C}$ ?

### SOLUTION

The probability of finding an atom of  $^{13}\text{C}$  in a molecule of cholesterol follows a binomial distribution, where  $X$  is the number of  $^{13}\text{C}$  atoms,  $N$  is the number of carbon atoms in a molecule of cholesterol, and  $p$  is the probability that any single atom of carbon is  $^{13}\text{C}$ .

- The mean number of  $^{13}\text{C}$  atoms in a molecule of cholesterol is

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

with a standard deviation of

$$\sigma = \sqrt{Np(1-p)} = \sqrt{27 \times 0.0111 \times (1-0.0111)} = 0.172$$

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

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

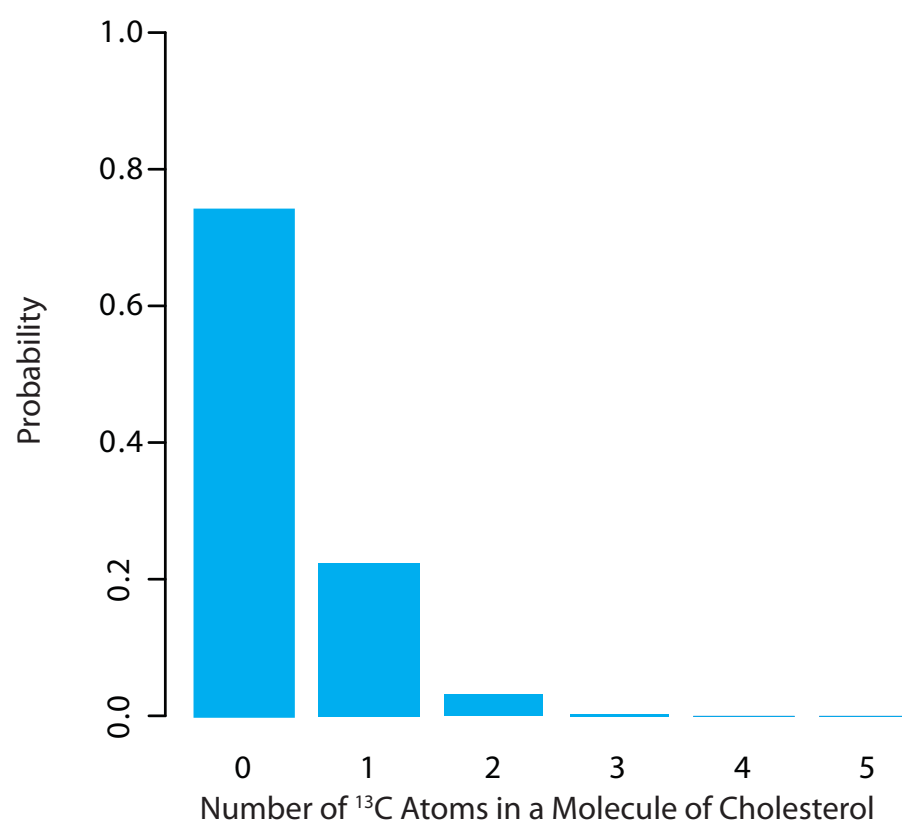
There is a 74.0% probability that a molecule of cholesterol will not have an atom of  $^{13}\text{C}$ , a result consistent with the observation that the mean number of  $^{13}\text{C}$  atoms per molecule of cholesterol, 0.300, is less than one.

A portion of the binomial distribution for atoms of  $^{13}\text{C}$  in cholesterol is shown in Figure 4.6. Note in particular that there is little probability of finding more than two atoms of  $^{13}\text{C}$  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  $^{13}\text{C}$  atoms in cholesterol. A molecule of cholesterol, for example, can have two  $^{13}\text{C}$  atoms, but it can not have 2.5 atoms of  $^{13}\text{C}$ . A population is continuous if its members may take on any value. The efficiency of extracting cholesterol from a sample, for example, can take on any value between 0% (no cholesterol extracted) and 100% (all cholesterol extracted).

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



**Figure 4.6** Portion of the binomial distribution for the number of naturally occurring  $^{13}\text{C}$  atoms in a molecule of cholesterol. Only 3.6% of cholesterol molecules contain more than one atom of  $^{13}\text{C}$ , and only 0.33% contain more than two atoms of  $^{13}\text{C}$ .



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

where  $\mu$  is the expected mean for a population with  $n$  members

$$\mu = \frac{\sum_i X_i}{n}$$

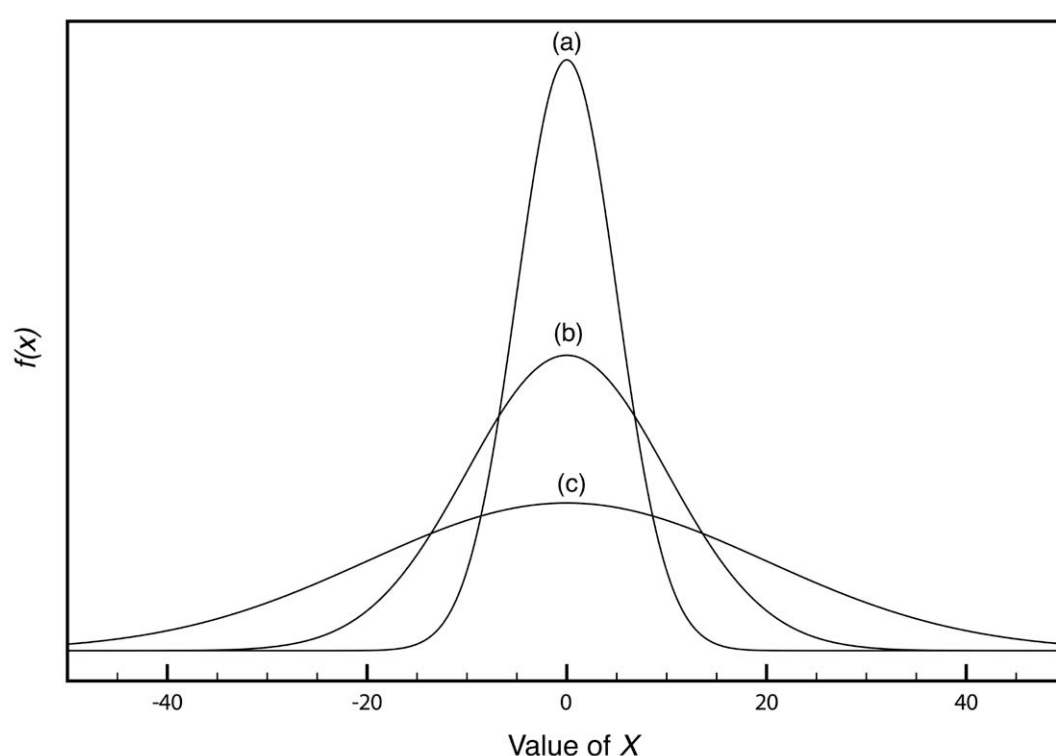
and  $\sigma^2$  is the population's variance.

$$\sigma^2 = \frac{\sum_i (X_i - \mu)^2}{n} \quad 4.8$$

Examples of normal distributions, each with an expected mean of 0 and with variances of 25, 100, or 400, 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 corresponding to  $\mu$ , 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 distribution.

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 with 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.

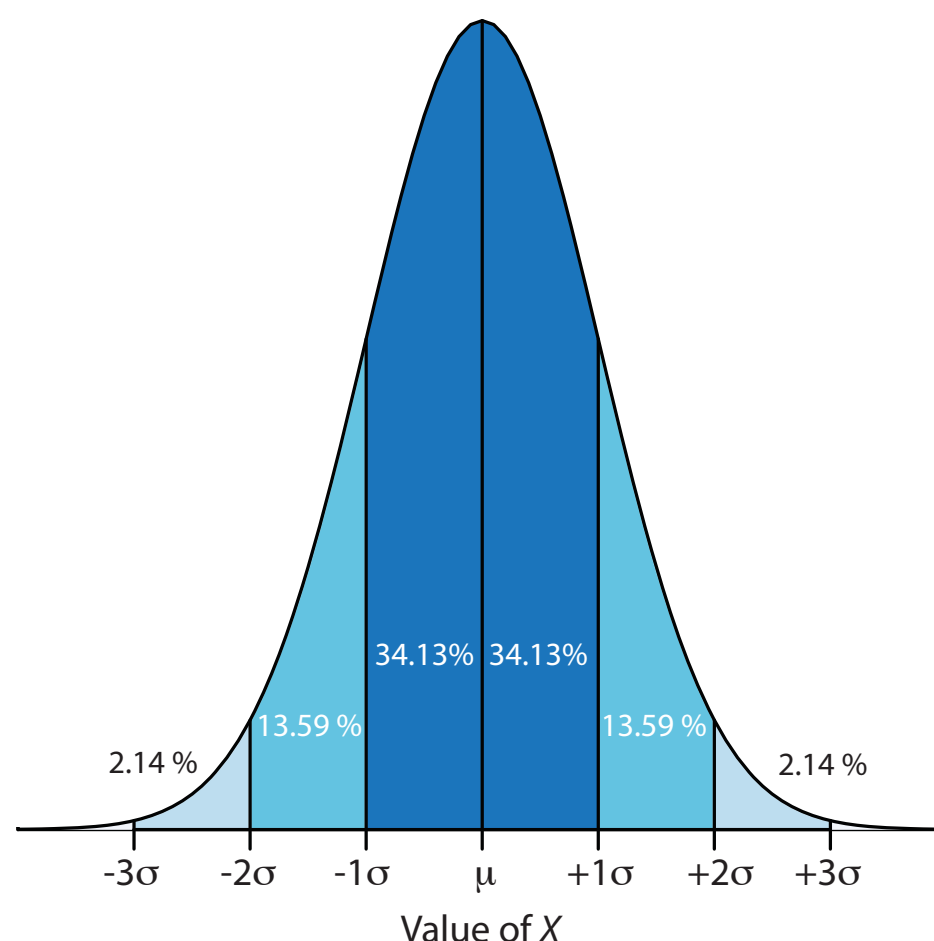
Because a normal distribution depends solely on  $\mu$  and  $\sigma^2$ , the probability of finding a member of the population between any two limits is the same for all normally distributed populations. Figure 4.8, for example, shows that 68.26% of the members of a normal distribution have a value



**Figure 4.7** Normal distribution curves for:

- (a)  $\mu = 0; \sigma^2 = 25$
- (b)  $\mu = 0; \sigma^2 = 100$
- (c)  $\mu = 0; \sigma^2 = 400$

**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 can be found using the probability table in [Appendix 3](#).



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 0.17% members of a population have values exceeding the expected mean by more than  $\pm 3\sigma$ . Additional ranges and probabilities are gathered together in a probability table that you will find in [Appendix 3](#). As shown in Example 4.11, if we know the mean and 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  $\sigma^2 = 25$ . In a random sampling 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 containing between 243 mg and 262 mg of aspirin.

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  $\mu$  in terms of the population's standard deviation,  $\sigma$

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

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

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

and the deviation for the upper limit is

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

Using the table in [Appendix 3](#), we find that the percentage of tablets with less than 243 mg of aspirin is 8.08%, and 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

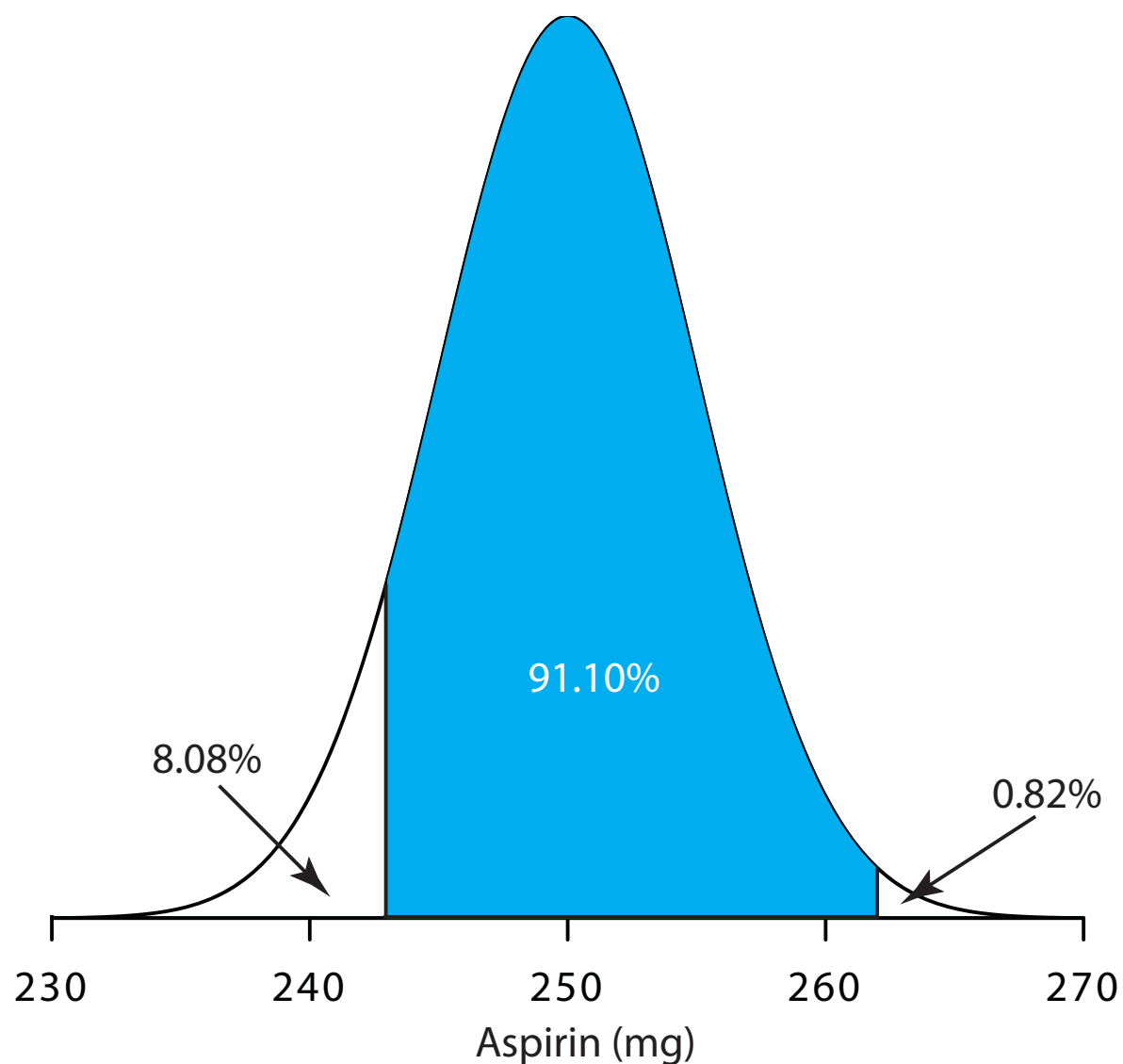
$$100.00\% - 8.08\% - 0.82\% = 91.10\%$$

Figure 4.9 shows the distribution of aspirin 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 randomly select 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 are found within the range of  $\mu \pm 1\sigma$ . 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,  $X_i$  to be in the range

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

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 we will discuss 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  $\mu$  is 250 mg and  $\sigma^2$  is 25?

#### 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  $250 \text{ mg} \pm 10 \text{ mg}$  means that 95% of the tablets in the population contain between 240 and 260 mg of aspirin.

Alternatively, we can express a confidence interval for the expected mean in terms of the population's standard deviation and the value of a single member drawn from the population.

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

### 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, there is 95% probability that the population's mean,  $\mu$ , 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. We can extend confidence intervals to include the mean of  $n$  samples drawn from a population of known  $\sigma$ . The standard deviation of the mean,  $\sigma_{\bar{x}}$ , which also is known as the **STANDARD ERROR OF THE MEAN**, is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

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

$$\mu = \bar{X} \pm \frac{z\sigma}{\sqrt{n}} \quad 4.11$$

[Problem 8](#) at the end of the chapter asks you to derive this equation using a propagation of uncertainty.

### Example 4.14

What is the 95% confidence interval for the analgesic tablets described 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

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

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

### 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 working example 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 situations where we can not study the whole population, or when we can not 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. To gain a better picture of the distribution of pennies we need a larger sample, such as that shown in [Table 4.13](#). The mean and the standard deviation for this sample of 100 pennies 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 mass intervals 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, assuming that the mean and variance for the sample provide appropriate estimates for the mean and variance of the population. Although the histogram is not perfectly symmetric, it provides a good approximation of the normal distribution curve, suggesting that the sample of 100 pennies

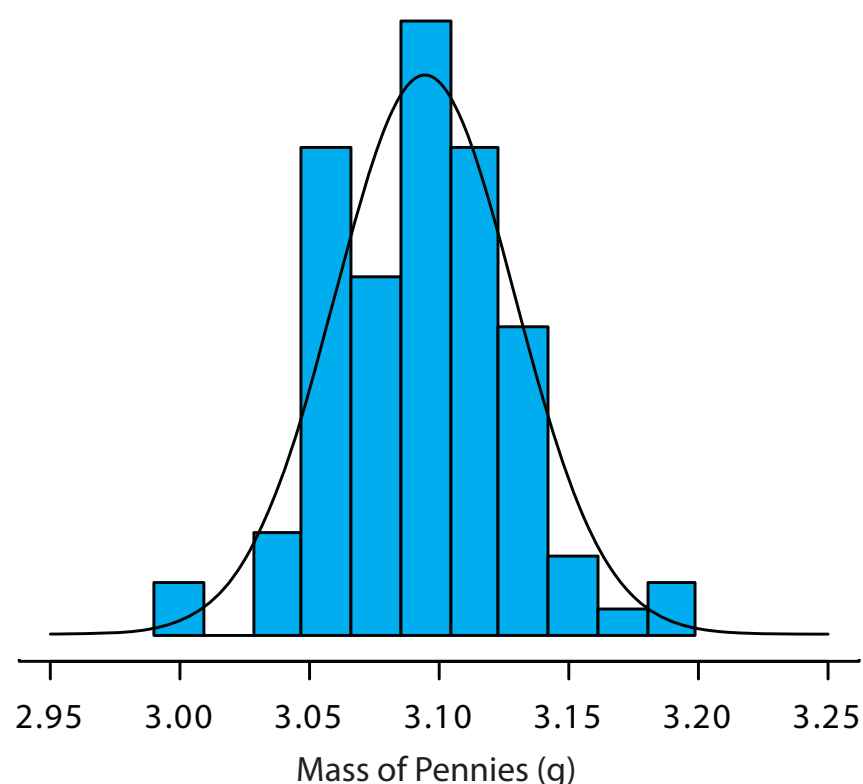
Table 4.13 Masses for a Sample of 100 Circulating U. S. Pennies

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	3.104–3.123	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	15	3.181–3.199	2
3.086–3.104	23		

**Figure 4.10** The blue bars show a histogram for the data in [Table 4.13](#). The height of a bar corresponds to the percentage of pennies within the mass intervals shown in [Table 4.14](#). Superimposed on the histogram is a normal distribution curve assuming that  $\mu$  and  $\sigma^2$  for the population are equivalent to  $\bar{X}$  and  $s^2$  for the sample. The total area of the histogram's bars and the area under the normal distribution curve are equal.



is normally distributed. It is easy to imagine that the histogram will more closely approximate a normal distribution if we include additional pennies in our sample.

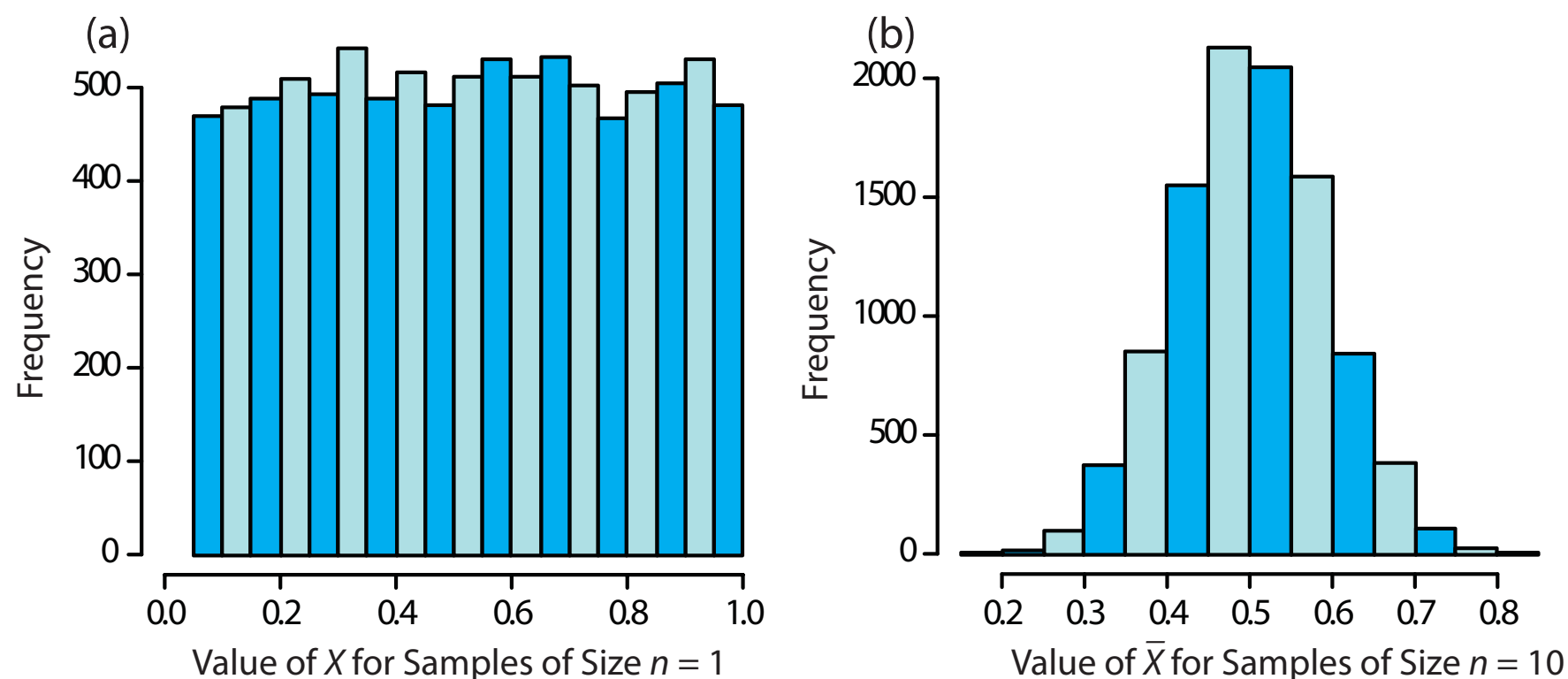
We will not offer a formal proof that the sample of pennies in [Table 4.13](#) and the population of all circulating U. S. pennies are normally distributed. The evidence we have seen, however, strongly suggests that this is true. Although we can not claim that the results for all analytical experiments are normally distributed, in most cases the data we collect in the laboratory are, in fact, drawn from a normally distributed population. According to the **CENTRAL LIMIT THEOREM**, when a system is subject to a variety of indeterminate errors, the results approximate a normal distribution.<sup>6</sup> As the number of sources of indeterminate error increases, the results more closely approximate a normal distribution. 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 closely approximate a normal distribution if the size of the samples is large enough. [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.

<sup>6</sup> Mark, H.; Workman, J. *Spectroscopy* **1988**, *3*, 44–48.

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 for analysis—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 the gross sample, therefore, is the mean for this large number of individual soil particles. Because of this, the central limit theorem is relevant.





**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.

### DEGREES OF FREEDOM

In reading to this point, did you notice the differences between the equations for the standard deviation or variance of a population and the standard deviation or variance of a sample? If not, here are the two equations:

$$\sigma^2 = \frac{\sum_i (X_i - \mu)^2}{n}$$

$$s^2 = \frac{\sum_i (X_i - \bar{X})^2}{n-1}$$

Both equations measure the variance around the mean, using  $\mu$  for a population and  $\bar{X}$  for a sample. Although the equations use different measures for the mean, the intention is the same for both the sample and the population. A more interesting difference is between the denominators of the two equations. In calculating the population's variance we divide the numerator by the population's size,  $n$ . For the sample's variance we divide by  $n-1$ , where  $n$  is the size of the sample. Why do we make this distinction?

A variance is the average squared deviation of individual results from the mean. In calculating an average we divide by the number of independent measurements, also known as the **DEGREES OF FREEDOM**, contributing to the calculation. For the population's variance, the degrees of freedom is equal to the total number of members,  $n$ , in the population. In measuring

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 there are  $n-1$  degrees of freedom remaining for calculating the sample's variance.

every member of the population we have complete information about the population.

When calculating the sample's variance, however, we first replace  $\mu$  with  $\bar{X}$ , which we also calculate from the same data. If there are  $n$  members in the sample, we can deduce the value of the  $n^{\text{th}}$  member from the remaining  $n-1$  members. 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

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

Using  $n-1$  in place of  $n$  when calculating the sample's variance ensures that  $s^2$  is an unbiased estimator of  $\sigma^2$ .

#### 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,  $\mu$ ,

$$\mu = \bar{X} \pm \frac{z\sigma}{\sqrt{n}} \quad 4.11$$

where  $\bar{X}$  is the mean for a sample of size  $n$ , and  $\sigma$  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,  $\sigma$ , with the sample's standard deviation,  $s$ . The second modification is less obvious. The values of  $z$  in [Table 4.12](#) are for a normal distribution, which is a function of  $\sigma^2$ , not  $s^2$ . Although the sample's variance,  $s^2$ , provides an unbiased estimate for the population's variance,  $\sigma^2$ , the value of  $s^2$  for any sample may differ significantly from  $\sigma^2$ . To account for the uncertainty in estimating  $\sigma^2$ , 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.

$$\mu = \bar{X} \pm \frac{ts}{\sqrt{n}} \quad 4.12$$

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, approaching  $z$  as  $n$  approaches infinity. The larger the sample, the more closely its confidence interval approaches the confidence interval given by 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	$\infty$	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 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 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, there is a 95% probability that the population's mean is between 3.070 to 3.164 g. For the second experiment, the 95% confidence interval spans 3.035 g–3.127 g. The two confidence intervals are not identical, but 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.

Our comparison of these two confidence intervals is rather 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 analyzing data 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 constantly reviewing and evaluating your data. For example, if an analysis on five samples gives 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 suggests that the analyte's true concentration lies within the range of  $-0.12$  ppm to 1.46 ppm. Including a negative concentration within the confidence interval should lead you to reevaluate your data or 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 detect accurately.

Here is a second example of why you should closely examine your data. The results for samples drawn 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 may not be normally distributed, or there may be a time-dependent determinate error. For example, if we randomly select 20 pennies and find that the mass of each penny is larger than that for the preceding penny, 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 indicates the range of values in which we expect to find the population's expected 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 claiming that there is only a 5% probability that the expected mass of penny is less than 3.070 g or more than 3.164 g.

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

Because a confidence interval is a statement of probability, it allows us to answer questions such as “Are the results for a newly developed method for determining 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.

#### 4E.1 Significance Testing

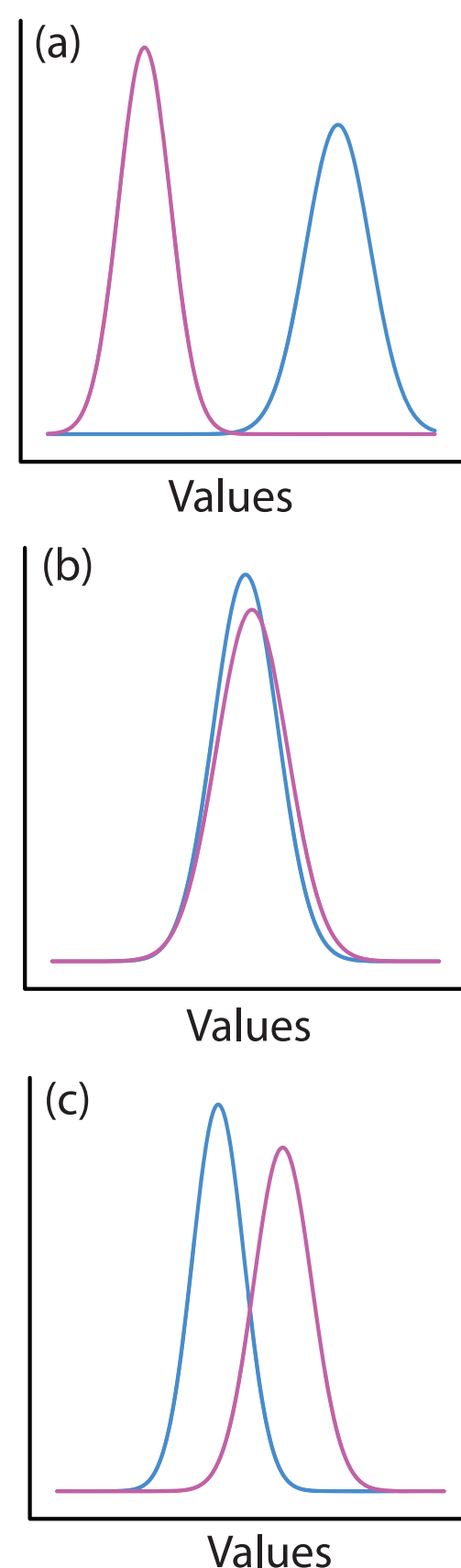
Let's consider the following problem. To determine if a medication is effective in lowering blood glucose concentrations, we collect two sets of blood samples from a patient. We collect one set of samples immediately before administering the medication, and collect the second set of samples several hours 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 normal distribution curves for each sample, and to compare them to each other. Three possible outcomes are shown in Figure 4.12. In Figure 4.12a, there is a complete separation of the normal distribution curves, strongly suggesting that the samples are significantly different. In Figure 4.12b, the normal distributions for the two samples almost completely overlap each other, suggesting that any difference between the samples is insignificant. Figure 4.12c, however, presents a dilemma. Although the means for the two samples appear to be different, there is sufficient overlap of the normal distributions 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.

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 discussing specific examples we will first establish a general approach to conducting and interpreting significance tests.

#### 4E.2 Constructing a Significance Test

The purpose of a **SIGNIFICANCE TEST** is to determine whether the difference between two or more values is too large to be explained by indeterminate error. 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 provide answers to the question. The **NULL HYPOTHESIS**,  $H_0$ , is that indeterminate error is sufficient to explain any differences in our data. The **ALTERNATIVE HYPOTHESIS**,  $H_A$ , is that the differences are too great to be explained by random error and, therefore, must be determinate. We test



**Figure 4.12** Three examples showing possible relationships between the normal distribution curves for two samples. In (a) the curves are completely separate, suggesting that the samples are significantly different from each other. In (b) the two curves are almost identical, suggesting that the samples are indistinguishable. The partial overlap of the curves in (c) means that the best we can do is to indicate the probability that there is a difference between the samples.

the null hypothesis, which we either retain or reject. If we reject the null hypothesis, then we must accept the alternative hypothesis, concluding that the difference is significant and that it cannot be explained by random error.

Failing to reject a null hypothesis is not the same as accepting it. We retain a null hypothesis because there is insufficient evidence to prove it incorrect. It is impossible to prove that a null hypothesis is true. This is an important point 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 you might propose the following null and alternative hypotheses.

$H_0$ : The mass of any U.S. penny in circulation is in the range of 2.900g–3.200 g.

$H_A$ : A U.S. penny in circulation may have a mass less than 2.900 g or a mass of more than 3.200 g.

To test the null hypothesis you reach into your pocket, retrieve a penny, and determine its mass. If the penny's mass is 2.512 g then you reject the null hypothesis, and accept the alternative hypothesis. Suppose that the penny's mass is 3.162 g. Although this result increases your confidence in the null hypothesis, it does not prove it is correct because the next penny you pull from your pocket might weigh less than 2.900 g or more than 3.200 g.

After stating the null and 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 in correctly rejecting the null hypothesis (e.g. 95%), or as the probability that we are incorrectly rejecting the null hypothesis. For the latter, the confidence level is given as  $\alpha$ , where

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

For a 95% confidence level,  $\alpha$  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 containing a known concentration of analyte,  $\mu$ . We analyze the standard several times, obtaining an mean value,  $\bar{X}$ , for the analyte's concentration. Our null hypothesis is that there is no difference between  $\bar{X}$  and  $\mu$

The four steps for a statistical analysis of data:

1. Pose a question, and state the null hypothesis and the alternative hypothesis.
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  $\alpha$  to represent the probability of incorrectly rejecting 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.

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

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

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

then we reject the null hypothesis and accept the alternative hypothesis if  $\mu$  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. This is a **TWO-TAILED SIGNIFICANCE TEST** because we reject the null hypothesis for values of  $\mu$  at either extreme of the sample's probability distribution curve.

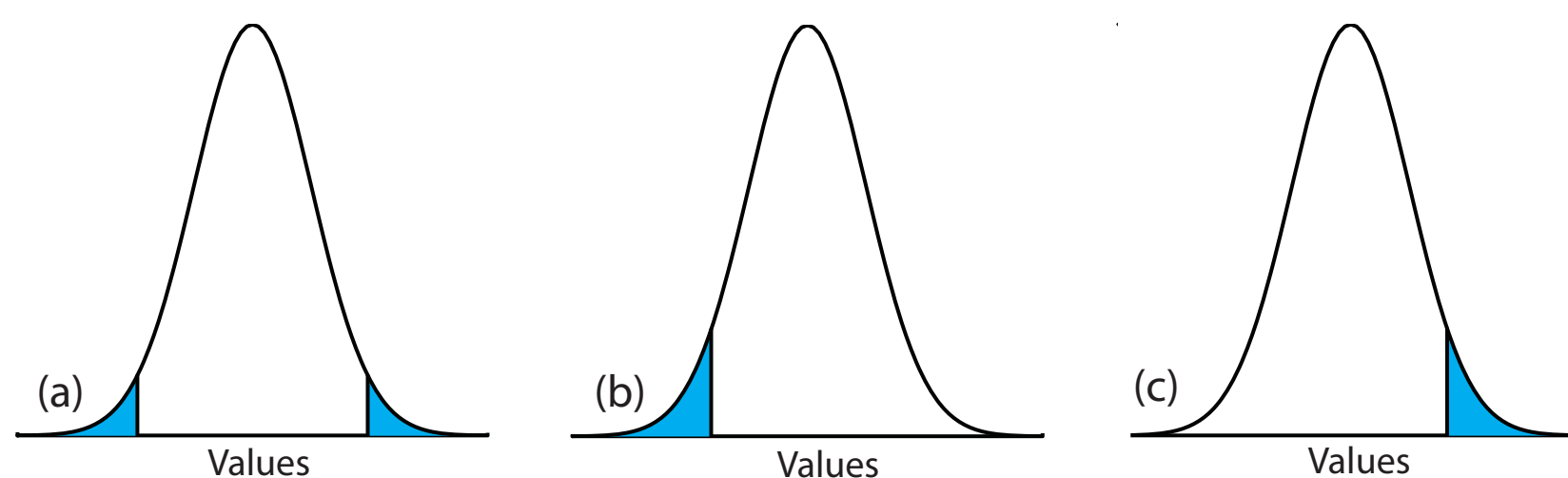
We also can write the alternative hypothesis in two additional ways

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

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

rejecting the null hypothesis if  $\mu$  falls within the shaded areas shown in Figure 4.13b or Figure 4.13c, respectively. In each case the shaded area 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 always a 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 smaller) than the other parameter. This is the case, for example, in evaluating the accuracy of a new analytical method. A two-tailed significance test, therefore, is usually the appropriate choice.



**Figure 4.13** Examples of a (a) two-tailed, and a (b, c) a one-tailed, significance test of  $\bar{X}$  and  $\mu$ . The normal distribution curves are drawn using 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  $\mu$  is within the blue areas, then we reject the null hypothesis and accept the alternative hypothesis. We retain the null hypothesis if the value of  $\mu$  is within the unshaded area of the curve.

We reserve a one-tailed significance test for a situation where we are specifically 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 result after administering the medication is less than the result before beginning treatment. If the patient's blood glucose level is greater after administering the medication, then we know the answer—the medication did not work—without conducting a statistical analysis.

#### 4E.4 Errors in Significance Testing

Because a significance test relies on probability, its interpretation is naturally subject to error. In a significance test,  $\alpha$  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 I ERROR**, and its risk is always equivalent to  $\alpha$ . Type 1 errors in two-tailed and one-tailed significance tests correspond 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 its probability of occurrence is  $\beta$ . Unfortunately, in most cases we cannot calculate or estimate the value for  $\beta$ . 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  $\alpha$  increases the likelihood of a type 2 error. When we choose a value for  $\alpha$  we are making a compromise between these two types of error. Most of the examples in this text use a 95% confidence level ( $\alpha = 0.05$ ) because this provides a reasonable compromise between type 1 and type 2 errors. It is not unusual, however, to use more stringent (e.g.  $\alpha = 0.01$ ) or more lenient (e.g.  $\alpha = 0.10$ ) confidence levels.

### 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 is well defined, constructing and evaluating significance tests is straightforward.

#### 4F.1 Comparing $\bar{X}$ to $\mu$

One approach for validating a new analytical method is to analyze a sample containing a known amount of analyte,  $\mu$ . To judge the method's accuracy we analyze several portions of the sample, determine the average amount of analyte in the sample,  $\bar{X}$ , and use a significance test to compare it to  $\mu$ . Our null hypothesis,  $H_0 : \bar{X} = \mu$ , is that any difference between  $\bar{X}$  and  $\mu$  is the result of indeterminate errors affecting the determination of  $\bar{X}$ . The



alternative hypothesis,  $H_A: \bar{X} \neq \mu$ , is that the difference between  $\bar{X}$  and  $\mu$  is too large to be explained by indeterminate error.

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

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

Rearranging this equation and solving for  $t_{\text{exp}}$

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

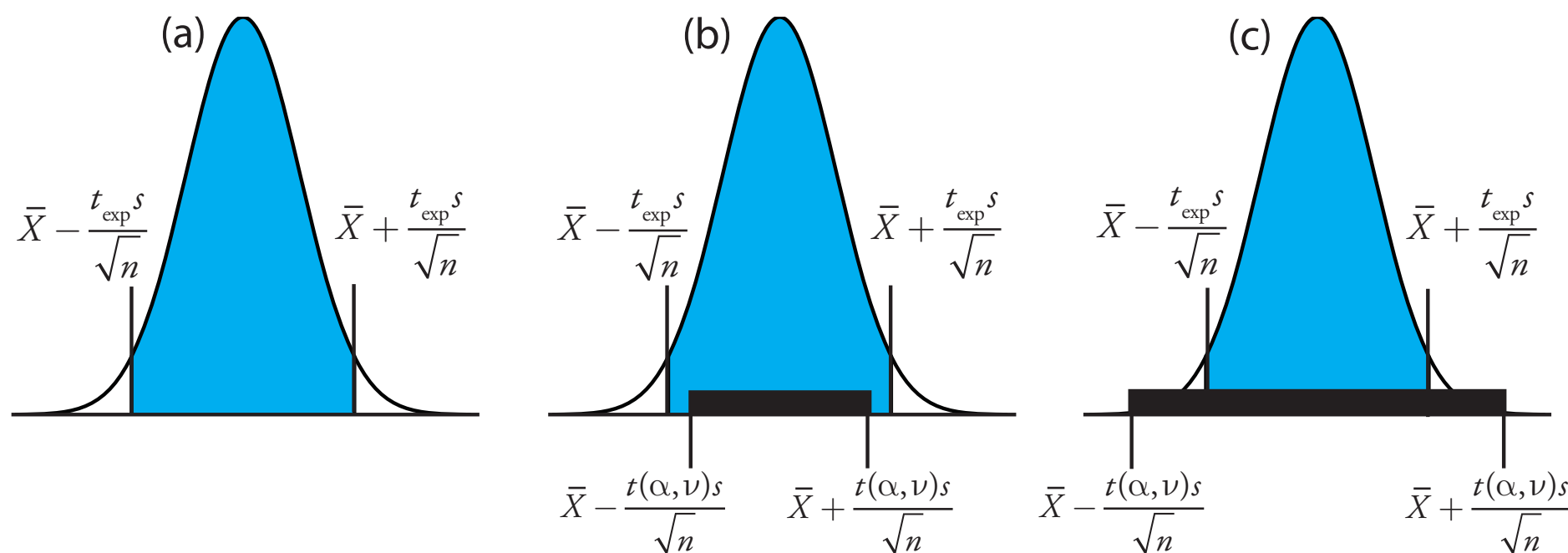
gives the value of  $t_{\text{exp}}$  when  $\mu$  is at either the right edge or the left edge of the sample's confidence interval (Figure 4.14a).

To determine if we should retain or reject the null hypothesis, we compare the value of  $t_{\text{exp}}$  to a critical value,  $t(\alpha, \nu)$ , where  $\alpha$  is the confidence level and  $\nu$  is the degrees of freedom for the sample. The critical value  $t(\alpha, \nu)$  defines the largest confidence interval resulting from indeterminate errors. If  $t_{\text{exp}} > t(\alpha, \nu)$ , then our sample's confidence interval is too large to be 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 the confidence interval for our sample can be explained by indeterminate 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  $\bar{X}$  to  $\mu$ .

Values for  $t(\alpha, \nu)$  are in [Appendix 4](#).

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 Guinness 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.



**Figure 4.14** Relationship between confidence intervals and the result of a significance test. (a) The shaded area under the normal distribution curve shows the confidence interval for the sample based on  $t_{\text{exp}}$ . Based on the sample, we expect  $\mu$  to fall within the shaded area. The solid bars in (b) and (c) show the confidence intervals for  $\mu$  that can be explained by indeterminate error given the choice of  $\alpha$  and the available degrees of freedom,  $\nu$ . For (b) we must reject the null hypothesis because there are portions of the sample's confidence interval that lie outside the confidence interval due to indeterminate error. In the case of (c) we retain the null hypothesis because the confidence interval due to indeterminate error completely encompasses the sample's confidence interval.

### Example 4.16

Before determining the amount of  $\text{Na}_2\text{CO}_3$  in a sample, you decide to check your procedure by analyzing a sample known to be 98.76% w/w  $\text{Na}_2\text{CO}_3$ . Five replicate determinations of the %w/w  $\text{Na}_2\text{CO}_3$  in the standard gave the following results.

98.71% 98.59% 98.62% 98.44% 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

$$\bar{X} = 98.59 \quad s = 0.0973$$

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

$$H_0 : \bar{X} = \mu \quad H_A : \bar{X} \neq \mu$$

The test statistic,  $t_{\text{exp}}$ , is

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

The critical value for  $t(0.05, 4)$  from [Appendix 4](#) is 2.78. Since  $t_{\text{exp}}$  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  $\bar{X}$  and  $\mu$  is too large to be explained by indeterminate sources of error, suggesting that a determinate source of error is affecting 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  $\mu$  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 you need to exercise caution when interpreting the results of a statistical analysis. We will keep returning to this point because it is an important one. Having determined that a result is

There is another way to interpret the result of this  $t$ -test. Knowing that  $t_{\text{exp}}$  is 3.91 and that there are 4 degrees of freedom, we use [Appendix 4](#) to estimate the  $\alpha$  value corresponding to a  $t(\alpha, 4)$  of 3.91. From [Appendix 4](#),  $t(0.02, 4)$  is 3.75 and  $t(0.01, 4)$  is 4.60. Although we can reject the null hypothesis at the 98% confidence level, we cannot reject it at the 99% confidence level.

For a discussion of the advantages of this approach, see J. A. C. Sterne and G. D. Smith "Sifting the evidence—what's wrong with significance tests?" [BMJ](#) **2001**, *322*, 226–231.

inaccurate, as we did in [Example 4.16](#), the next step is to identify and to correct the error. Before expending time and money on this, however, you should first critically examine your data. For example, the smaller the value of  $s$ , the larger the value of  $t_{\text{exp}}$ . If the standard deviation for your analysis is unrealistically small, the probability of a type 2 error increases. Including a few additional replicate analyses of the standard and reevaluating the  $t$ -test may strengthen your evidence for a determinate error, or it may show that there is no evidence for a determinate error.

## 4F.2 Comparing $s^2$ to $\sigma^2$

If we regularly analyze a particular sample, we may be able to establish the expected variance,  $\sigma^2$ , for the analysis. This often is the case, for example, in clinical labs that routinely 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  $\sigma^2$ .

We can use an **F-TEST** to evaluate whether a difference between  $s^2$  and  $\sigma^2$  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  $F_{\text{exp}}$ , which is given as either

$$F_{\text{exp}} = \frac{s^2}{\sigma^2} \quad \text{or} \quad F_{\text{exp}} = \frac{\sigma^2}{s^2} \quad 4.16$$

$$(s^2 > \sigma^2) \quad (s^2 < \sigma^2)$$

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

If the null hypothesis is true, then  $F_{\text{exp}}$  should equal one. Because of indeterminate errors, however,  $F_{\text{exp}}$  usually is greater than one. A critical value,  $F(\alpha, \nu_{\text{num}}, \nu_{\text{den}})$ , gives the largest value of  $F_{\text{exp}}$  that we can attribute to indeterminate error. It is chosen for a specified significance level,  $\alpha$ , and for the degrees of freedom for the variance in the numerator,  $\nu_{\text{num}}$ , and the variance in the denominator,  $\nu_{\text{den}}$ . 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  $\sigma^2$  is  $\infty$ . 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.

254 249 252 252 249 249 250 247 251 252

Determine whether there is any evidence of a significant difference between that the sample's variance 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

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

The value for  $F_{\text{exp}}$  is

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

The critical value for  $F(0.05, \infty, 9)$  from [Appendix 5](#) is 3.333. Since  $F_{\text{exp}}$  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

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

defining A and B so that the value of  $F_{\text{exp}}$  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 precisions 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

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

and the value of  $F_{\text{exp}}$  is

$$F_{\text{exp}} = \frac{s_A^2}{s_B^2} = \frac{0.00259}{0.00138} = 1.88$$

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

### Practice Exercise 4.9

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?

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 of the factors while holding the others constant. 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 analysts or to compare 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 the 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.

- Example 1: 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: Collect 10 pennies and divide them into two groups of five pennies each. Weigh the pennies in the first group on one balance and 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 pennies are from the same population. The difference is how we sample the 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  $\bar{X}_A$  and  $\bar{X}_B$ , and standard deviations of  $s_A$  and  $s_B$ . The confidence intervals for  $\mu_A$  and for  $\mu_B$  are

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](#).

One simple test for determining whether data are paired or unpaired is to look at the size of each sample. If the samples are of different size, then the data must be unpaired. The converse is not true. If two samples are of equal size, they may be paired or unpaired.

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

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

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

To derive an equation for  $t_{\text{exp}}$ , we assume that  $\mu_A$  equals  $\mu_B$ , and combine equations 4.17 and 4.18.

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

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

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

Finally, we solve for  $t_{\text{exp}}$

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

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

Thus far our development of this  $t$ -test is similar to that for comparing  $\bar{X}$  to  $\mu$ , 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  $\sigma^2$ . In this case we can replace  $s_A^2$  and  $s_B^2$  with a pooled variance,  $(s_{\text{pool}})^2$ , that provides a better estimate for the variance. Thus, equation 4.20 becomes

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

where  $s_{\text{pool}}$ , the pooled standard deviation, is

[Problem 9](#) asks you to use a propagation of uncertainty to show that equation 4.19 is correct.

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

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

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.

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

$$v = \frac{\left[ \frac{s_A^2}{n_A} + \frac{s_B^2}{n_B} \right]^2}{\frac{\left( \frac{s_A^2}{n_A} \right)^2}{n_A + 1} + \frac{\left( \frac{s_B^2}{n_B} \right)^2}{n_B + 1}} - 2 \quad 4.23$$

Since the degrees of freedom must be an integer, we round to the nearest integer the value of  $v$  obtained using equation 4.23.

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

### 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 must determine whether we can pool the variances. This is done using an  $F$ -test. We did this analysis in [Example 4.18](#), finding no evidence of a significant difference. The pooled standard deviation is

$$s_{\text{pool}} = \sqrt{\frac{(7-1)(0.00259)^2 + (5-1)(0.00138)^2}{7+5-2}} = 0.0459$$

with 10 degrees of freedom. To compare the means the null hypothesis and alternative hypotheses are

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

Because we are using the pooled standard deviation, we calculate  $t_{\text{exp}}$  using [equation 4.21](#).

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

The critical value for  $t(0.05, 10)$ , from [Appendix 4](#), is 2.23. Because  $t_{\text{exp}}$  is less than  $t(0.05, 10)$  we retain the null hypothesis. For  $\alpha = 0.05$  there is no evidence that the two sets of pennies are significantly different.

### Example 4.20

One method for determining the %w/w  $\text{Na}_2\text{CO}_3$  in soda ash is 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	83.94

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

#### SOLUTION

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

$$\begin{aligned} \bar{X}_A &= 86.83\% & \bar{X}_B &= 82.71\% \\ s_A &= 0.32 & s_B &= 2.16 \end{aligned}$$

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

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

Calculating  $F_{\text{exp}}$ , we obtain a value of

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

Because  $F_{\text{exp}}$  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. As a result, we cannot calculate a pooled standard deviation.



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

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

Because we cannot pool the standard deviations, we calculate  $t_{\text{exp}}$  using [equation 4.20](#) instead of [equation 4.21](#)

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

and calculate the degrees of freedom using [equation 4.23](#).

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

From [Appendix 4](#), the critical value for  $t(0.05,5)$  is 2.57. Because  $t_{\text{exp}}$  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

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.

difference between the methods if we use different patients to gather data for each method. Using paired data, in which we analyze each patient's blood using both methods, prevents a large variance within a population from adversely affecting a  $t$ -test of means.

When using 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,  $\bar{d}$ , and the standard deviation of the differences,  $s_d$ . The null hypothesis,  $H_0 : \bar{d} = 0$ , is that there is no difference between the two samples, and the alternative hypothesis,  $H_A : \bar{d} \neq 0$ , is that the difference between the two samples is significant.

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

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

where  $n$  is the number of paired samples. As is true for other forms of the  $t$ -test, we compare  $t_{\text{exp}}$  to  $t(\alpha, \nu)$ , where the degrees of freedom,  $\nu$ , is  $n - 1$ . If  $t_{\text{exp}}$  is greater than  $t(\alpha, \nu)$ , then we reject the null hypothesis and accept the alternative hypothesis. We retain the null hypothesis if  $t_{\text{exp}}$  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 rapidly determining the concentration of the antibiotic monensin in fermentation vats.<sup>7</sup> The standard method for the analysis, a test for microbiological activity, is both difficult 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

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

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

$$H_0 : \bar{d} = 0 \quad H_A : \bar{d} \neq 0$$

Defining the difference between the methods as

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

we can calculate the following differences for the samples.

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

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

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  $\alpha = 0.05$ .

### 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  $\text{Zn}^{2+}$  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 the data in the table shown to the right, determine if there is a significant difference between the concentration of  $\text{Zn}^{2+}$  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 (b) paired. Briefly comment on your results.

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

One important requirement for a paired  $t$ -test is that determinate and indeterminate errors affecting the analysis must be independent of the ana-

lyte'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  $\bar{d}$  and  $s_d$  leads to a biased estimate of the expected mean and standard deviation. This is rarely a problem for samples spanning 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 concentration. In such cases a paired  $t$ -test may give misleading results since 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.

Data that are not consistent with the remaining data are called outliers. An **OUTLIER** might exist for many reasons: the outlier might be from a different population (Is this a Canadian penny?); the outlier might be a contaminated or otherwise altered sample (Is the penny damaged?); 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 for identifying potential outliers, three of which we present here.

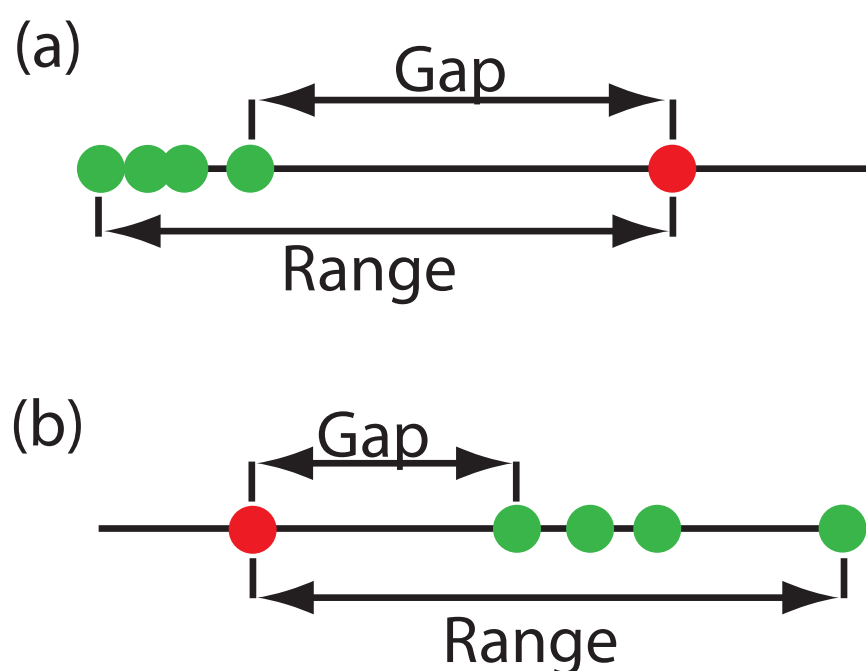
#### DIXON'S Q-TEST

One of the most common significance tests for outliers 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,  $Q_{\text{exp}}$ , is

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

**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



**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.

This equation is appropriate for evaluating a single outlier. Other forms of Dixon's  $Q$ -test allow its extension to detecting multiple outliers.<sup>8</sup>

The value of  $Q_{\text{exp}}$  is compared to a critical value,  $Q(\alpha, n)$ , where  $\alpha$  is the probability of rejecting 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, we usually apply the more conservative two-tailed  $Q$ -test, even though the possible outlier is the smallest or the largest value in the data set. If  $Q_{\text{exp}}$  is greater than  $Q(\alpha, n)$ , then we reject the null hypothesis and may exclude the outlier. We retain the possible outlier when  $Q_{\text{exp}}$  is less than or equal to  $Q(\alpha, n)$ . Table 4.17 provides values for  $Q(0.05, n)$  for a sample containing 3–10 values. A more extensive table is in [Appendix 6](#). Values for  $Q(\alpha, n)$  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

### 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 now recommends **GRUBB'S TEST**.<sup>9</sup> 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,  $G_{\text{exp}}$ , is the distance between the sample's mean,  $\bar{X}$ , and the potential outlier,  $X_{\text{out}}$ , in terms of the sample's standard deviation,  $s$ .

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

We compare the value of  $G_{\text{exp}}$  to a critical value  $G(\alpha, n)$ , where  $\alpha$  is the probability of rejecting a valid data point and  $n$  is the number of data points in the sample. If  $G_{\text{exp}}$  is greater than  $G(\alpha, n)$ , we may reject the data point as an

<sup>8</sup> Rorabacher, D. B. *Anal. Chem.* **1991**, *63*, 139–146.

<sup>9</sup> 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.

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

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.

### CHAUVENET'S CRITERION

Our final method for identifying outliers 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,  $z$

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

where  $X_{\text{out}}$  is the potential outlier,  $\bar{X}$  is the sample's mean and  $s$  is the sample's standard deviation. Note that this equation is identical to the equation for  $G_{\text{exp}}$  in the Grubb's test. For a normal distribution, you 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 of the values, 2.514 g, appears to be an outlier. Determine if this penny is an outlier using the 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  $Q_{\text{exp}}$  is

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

From [Table 4.17](#), the critical value for  $Q(0.05, 9)$  is 0.493. Because  $Q_{\text{exp}}$  is greater than  $Q(0.05, 9)$ , we can assume that penny weighing 2.514 g 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  $G_{\text{exp}}$  is

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

Using [Table 4.18](#), we find that the critical value for  $G(0.05,9)$  is 2.215. Because  $G_{\text{exp}}$  is greater than  $G(0.05,9)$ , we can assume that the penny weighing 2.514 g 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  $G_{\text{exp}}$ , 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 that the penny weighing 2.514 g 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 unreasonably better than that expected based on a propagation of uncertainty. Given these two concerns it is not surprising that some statisticians caution against the removal of outliers.<sup>10</sup>

On the other hand, testing for outliers can provide useful information if you 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 was changed from a brass alloy consisting of 95% w/w Cu and 5% w/w Zn, to a zinc core covered with copper.<sup>11</sup> 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.<sup>12</sup> 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  $S_{\text{mb}}$  represent the average signal for a method blank, and letting  $\sigma_{\text{mb}}$  represent the method blank's standard deviation. The null hypothesis is that the analyte is not present in the sample, and the alterna-

You also can adopt a more stringent requirement for rejecting data. When using the Grubb's test, for example, the ISO 5752 guidelines suggest 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 data point.

See [Chapter 3](#) for a review of these equations.

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>

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

$$(S_A)_{DL} = S_{mb} + \frac{ts_{mb}}{\sqrt{n}}$$

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.

tive 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,

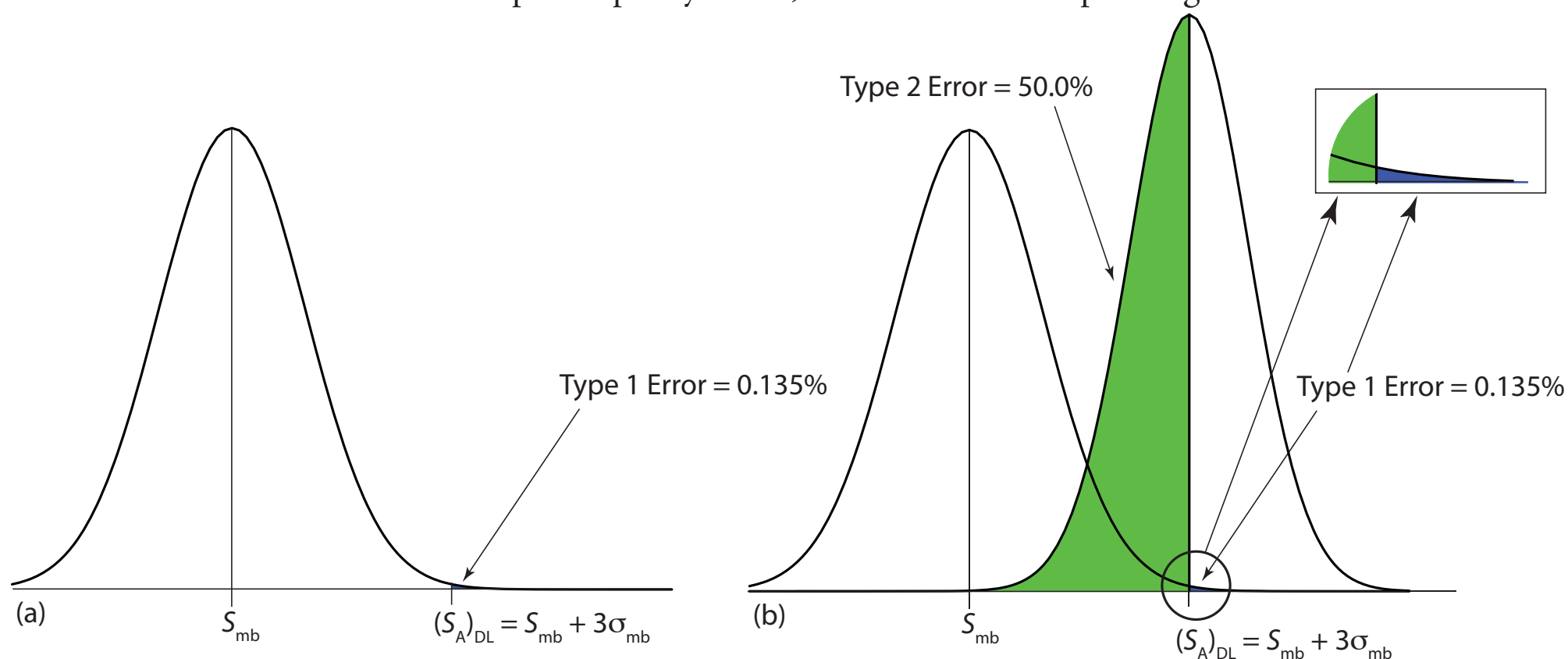
$$(S_A)_{DL} = S_{mb} + z\sigma_{mb} \quad 4.24$$

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

The value we choose for  $z$  depends on our tolerance for reporting the analyte's concentration even though it is absent from the sample (a type 1 error). Typically,  $z$  is set to three, which, from [Appendix 3](#), corresponds to a probability,  $\alpha$ , 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.

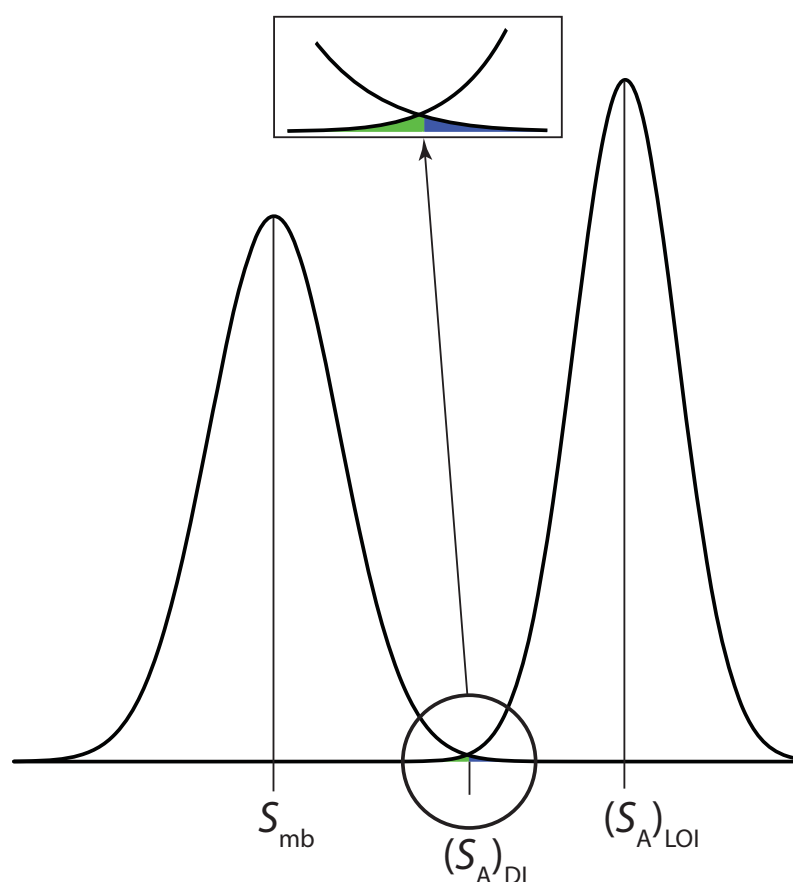
A detection limit also is subject to a type 2 error in which we fail to find evidence for the analyte even though it is present in the sample. Consider, for example, the situation shown in Figure 4.16b where the signal for a sample containing the analyte is exactly equal to  $(S_A)_{DL}$ . In this case the probability of a type 2 error is 50% because half of the signals arising from such samples are below the detection limit. We will correctly detect the analyte at the IUPAC detection limit only half the time. The IUPAC definition for the detection limit indicates the smallest signal for which we can say, at a significance level of  $\alpha$ , that an analyte is present in the sample. Failing to detect the analyte does not imply that it is not present in the sample.

The detection limit is often represented, particularly when discussing public policy issues, as a distinct line separating detectable concentrations



**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,  $\sigma_A$ , 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%.





**Figure 4.17** Normal distribution curves for a method blank and for a sample at the limit of identification:  $S_{\text{mb}} = 0$ ;  $\sigma_{\text{mb}} = 1$ ;  $\sigma_{\text{A}} = 0.8$ ; and  $(S_{\text{A}})_{\text{LOI}} = 0 + 3 \times 1 + 3 \times 0.8 = 5.4$ . 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%).

of analytes that concentrations that cannot be detected. This use of a detection limit is incorrect.<sup>13</sup> As suggested by [Figure 4.16](#), for concentrations of analyte near the detection limit there is a high probability of failing to detect the analyte.

An alternative expression for the detection limit, the **LIMIT OF IDENTIFICATION**, minimizes both type 1 and type 2 errors.<sup>14</sup> The analyte's signal at the limit of identification,  $(S_{\text{A}})_{\text{LOI}}$ , includes an additional term,  $z\sigma_{\text{A}}$ , to account for the distribution of the analyte's signal.

$$(S_{\text{A}})_{\text{LOI}} = (S_{\text{A}})_{\text{DL}} + z\sigma_{\text{A}} = S_{\text{mb}} + z\sigma_{\text{mb}} + z\sigma_{\text{A}}$$

As shown in [Figure 4.17](#), the limit of identification provides an equal probability for type 1 and type 2 errors 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 will be 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**,  $(S_{\text{A}})_{\text{LOQ}}$ .<sup>15</sup>

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

<sup>13</sup> Rogers, L. B. *J. Chem. Educ.* **1986**, *63*, 3–6.

<sup>14</sup> Long, G. L.; Winefordner, J. D. *Anal. Chem.* **1983**, *55*, 712A–724A.

<sup>15</sup> “Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry,” *Anal. Chem.* **1980**, *52*, 2242–2249.

## 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 descriptive statistics, for finding probabilities for different distributions, and for carrying out significance tests. In addition, R provides useful functions for visualizing your data.

### 4H.1 Excel

Excel provides two methods for working with data: built-in functions for individual statistical parameters and a package of data analysis tools in the Analysis ToolPak. The ToolPak is not a standard part of Excel's installation. 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.

#### 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. Select **Data Analysis...** from the **Tools** menu, which opens a window entitled "*Data Analysis*." Scroll through the window, select **Descriptive Statistics** from the available options, and click **OK**. Place the cursor in the box for the "*Input Range*" and then click and drag over the cells B1:B8. Check the box for "*Labels in the first row*." Select the radio button for "*Output range*," place the cursor in the box and click on an empty cell; this is where Excel will place the results. Check the boxes for "*Summary statistics*" and for the "*Confidence level for the mean*." Accept the default value of 95% for the confidence level. Clicking **OK** generates the information shown in Figure 4.19.

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

	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](#).

**Figure 4.19** Output from Excel's Descriptive Statistics command in the Analysis ToolPak. Note that Excel does not adjust for significant figures. The mode is the most common result, which is not relevant here. Kurtosis is a measure of the "peakedness" of the data's distribution, and is zero for a normal distribution. Skewness is a measure of the symmetry of the data's distribution and is zero for a normal distribution. For a small sample size—such as the seven samples in this data set—skewness and kurtosis are not particularly useful. You may consult the textbooks listed in the Additional Resources for more information about kurtosis and skewness.

<i>mass (g)</i>	
Mean	3.11728571
Standard Error	0.01924369
Median	3.107
Mode	#N/A
Standard Deviation	0.05091403
Sample Variance	0.00259224
Kurtosis	-0.59879248
Skewness	0.72905145
Range	0.142
Minimum	3.056
Maximum	3.198
Sum	21.821
Count	7
Confidence Level(95.0%)	0.04708762

Table 4.19 Excel Functions for Descriptive Statistics

Parameter	Excel Function
mean	=average(number1,number2,...)
median	=median(number1,number2,...)
sample standard deviation	=stdev(number1,number2,...)
population standard deviation	=stdevp(number1,number2,...)
sample variance	=var(number1,number2,...)
population variance	=varp(number1,number2,...)
range	=max((number1,number2,...) – min(number1,number2,...))

The **Descriptive Statistics** command provides a table of values for a sample. If your interest is in a statistical summary for a population, or if you do not want a summary of all descriptive statistics, you will need to use Excel's built-in functions (Table 4.19). To use a function, click on an empty cell, enter the formula, and press Return or Enter. To find, for example, the population variance for the data in [Figure 4.18](#), place the cursor in an empty cell and enter the formula

$$=varp(b2:b8)$$

The contents of the cell are replaced with Excel's exact calculation of the population's variance (0.002 221 918).

### 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,  $\mu$ , of 250 mg in terms of the population's expected standard deviation,  $\sigma$ , of 5 mg. After calculating values for  $z$ , we used the table in [Appendix 3](#) to find the area under the normal distribution curve between the two limits.

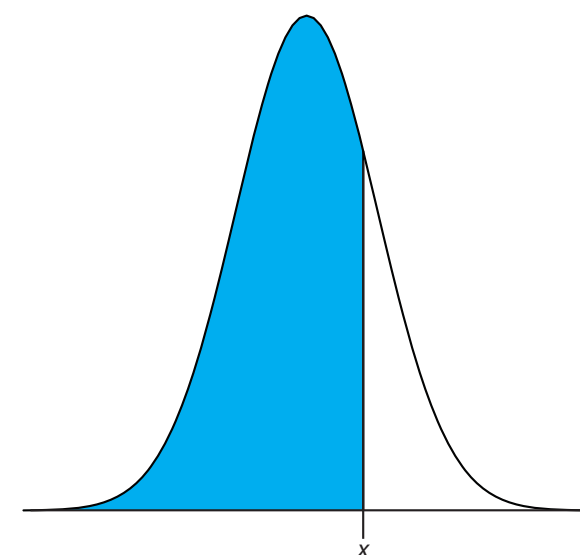
We can complete this calculation in Excel using the built-in **normdist** function. The function's general format is

$$=normdist(x, \mu, \sigma, TRUE)$$

where  $x$  is the limit of interest. The function returns the probability of obtaining a result of less than  $x$  from a normal distribution with a mean of  $\mu$  and a standard deviation of  $\sigma$  (Figure 4.20). To solve [Example 4.11](#) using Excel enter the following formulas into separate cells

$$=normdist(243, 250, 5, TRUE)$$

$$=normdist(262, 250, 5, TRUE)$$



**Figure 4.20** Shown in blue is the area returned by the function

$$=normdist(x, \mu, \sigma, TRUE)$$

obtaining results of 0.080756659 and 0.991802464. Subtract the smaller value from the larger value and adjust to the correct number of significant figures to arrive at a probability of 0.9910, or 99.10%.

Excel also includes a function for binomial distributions. The function's format is

$$=\text{binomdist}(X, N, p, T/F)$$

where  $X$  is the number of times a particular outcome occurs in  $N$  trials, and  $p$  is the probability of  $X$  occurring in one trial. Setting the function's last term to TRUE gives the total probability for any result up to  $X$  and setting it to FALSE to give the probability for  $X$ . Using [Example 4.10](#) to test this function, we find that the probability of finding no atoms of  $^{13}\text{C}$  atoms in a molecule of cholesterol,  $\text{C}_{27}\text{H}_{44}\text{O}$  using the formula

$$=\text{binomdist}(0, 27, 0.0111, \text{FALSE})$$

which returns a value of 0.740 as an answer, after adjusting the significant figures. Using the formula

$$=\text{binomdist}(2, 27, 0.0111, \text{TRUE})$$

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

### SIGNIFICANCE TESTS

Excel's Analysis ToolPak includes tools to help you complete 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 the ToolPak 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 as shown in Figure 4.21. Because the data in this case are unpaired, we will use Excel to complete an unpaired  $t$ -test. Before we can complete a  $t$ -test we must use an  $F$ -test to determine whether the variances for the two data sets are equal or unequal. Our null hypothesis is that the variances are equal,  $s_{\text{Set 1}}^2 = s_{\text{Set 2}}^2$ , and our alternative hypothesis is that the variances are not equal,  $s_{\text{Set 1}}^2 \neq s_{\text{Set 2}}^2$ .

To complete the  $F$ -test select **Data Analysis...** from the **Tools** menu, which opens a window entitled "*Data Analysis.*" Scroll through the window, select **F-Test Two Sample Variance** from the available options, and click **OK**. Place the cursor in the box for the "*Variable 1 range*" and then click and drag over the cells B1:B8. Next, place the cursor in the box for "*Variable 2 range*" and then click and drag over the cells B1:B6. Check the box

	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 4.21** Portion of a spreadsheet containing the data in [Table 4.11](#).

## F-Test Two-Sample for Variances

	Data Set 1	Data Set 2
Mean	3.11728571	3.0814
Variance	0.00259224	0.0013843
Observations	7	5
df	6	4
F	1.87259849	
P(F<=f) one-tail	0.28305251	
F Critical one-tail	9.19731108	

**Figure 4.22** Output from Excel's  $F$ -test in the Analysis ToolPak. Note that results are for a one-tailed  $F$ -test. The value for "P(F<=f) one-tail" is the probability of incorrectly rejecting the null hypothesis of equal variances, and the value for "F Critical one-tail" is equivalent to a one-tailed  $F(0.025, 6, 4)$ . For a two-tailed  $F$ -test, the probability of incorrectly rejecting the null hypothesis is  $2 \times P(F<=f)$ , or 0.566, and F Critical is for  $F(0.05, 6, 4)$ . *Note: Excel assumes that the variance for variable 1 is greater than the variance for variable 2; thus, you must assign variable 1 to the data set with the larger of the two variances in order to correctly interpret the  $F$ -test.*

for "Labels." In the box for "Alpha" enter your value for  $\alpha$ . Excel's  $F$ -test is a one-tailed test, so if you wish to complete a two-tailed test, as we do in this example, divide your  $\alpha$  value by 2. We want our  $F$ -test to use an  $\alpha$  of 0.05, so we enter 0.025 in the box for "Alpha." Select the radio button for "Output range," place the cursor in the box and click on an empty cell; this is where Excel will place the results. Clicking **OK** generates the information shown in Figure 4.22. Because  $F_{\text{exp}}$ , which is 1.87, is less than  $F(0.05, 6, 4)$ , which is 9.20, we retain the null hypothesis and have no evidence of a difference between the variances at an  $\alpha$  of 0.05.

Having found no evidence suggesting unequal variances, we now complete an unpaired  $t$ -test assuming equal variances. Our null hypothesis is that there is no difference between the means,  $\bar{X}_{\text{Set 1}} = \bar{X}_{\text{Set 2}}$ , and our alternative hypothesis is that there is a difference between the means,  $\bar{X}_{\text{Set 1}} \neq \bar{X}_{\text{Set 2}}$ . To complete the  $t$ -test select **Data Analysis...** from the **Tools** menu, which opens a window entitled "Data Analysis." Scroll through the window, select  **$t$ -Test Two Sample Assuming Equal Variances** from the available options, and click **OK**. Place the cursor in the box for the "Variable 1 range" and then click and drag over the cells B1:B8. Next, place the cursor in the box for "Variable 2 range" and then click and drag over the cells B1:B6. In the box for "Hypothesized mean difference," enter a value of 0. Check the box for "Labels." In the box for "Alpha" enter your value for  $\alpha$ . Select the radio button for "Output range," place the cursor in the box and click on an empty cell; this is where Excel will place the results. Clicking **OK** generates the information shown in Figure 4.23. Because  $t_{\text{exp}}$ , which is 1.33, is less than  $t(0.05, 10)$ , which is 2.23, we retain the null hypothesis and have no evidence of a difference between the means at an  $\alpha$  of 0.05.

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

See [Example 4.18](#) for our earlier solution to this problem.

See [Example 4.19](#) for our earlier solution to this problem.

t-Test: Two-Sample Assuming Equal Variances		
	Data Set 1	Data Set 2
Mean	3.11728571	3.0814
Variance	0.00259224	0.0013843
Observations	7	5
Pooled Variance	0.00210906	
Hypothesized Mean Difference	0	
df	10	
t Stat	1.33450508	
P(T<=t) one-tail	0.10581382	
t Critical one-tail	1.8124611	
P(T<=t) two-tail	0.21162765	
t Critical two-tail	2.22813884	

**Figure 4.23** Output from Excel's *t*-test assuming equal variances in the Analysis ToolPak. The absolute value of "t Stat" is equivalent to  $t_{\text{exp}}$  (note: because Excel subtracts the mean for variable 2 from the mean for variable 1, *t* Stat may be positive or negative). The values of "P(T<=t) one-tail" and of "P(T<=t) two-tail" are the probabilities of incorrectly rejecting the null hypothesis for a one-tailed or for a two-tailed *t*-test. The values for "t Critical one-tail" and for "t Critical two tail" are  $t(0.05, 10)$  for a one-tailed and two-tailed *t*-test, respectively.

### Practice Exercise 4.12

Rework [Example 4.20](#) and [Example 4.21](#) using Excel's Analysis ToolPak.

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

## 4H.2 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](http://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 containing the data, which we do by typing in the following command.

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

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

```
> var(penny1)
[1] 0.002221918
```

In R, the symbol '>' is a prompt indicating that the program is waiting for you to enter a command. When you press 'Return' or 'Enter,' R executes the 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)

### 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,  $\mu$ , of 250 mg in terms of the population's expected standard deviation,  $\sigma$ , of 5 mg. After calculating 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

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

where  $x$  is the limit of interest,  $\mu$  is the distribution's expected mean and  $\sigma$  is the distribution's expected standard deviation. The function returns the probability of obtaining a result of less than  $x$  (Figure 4.24). Here is the output of an R session for solving [Example 4.11](#).

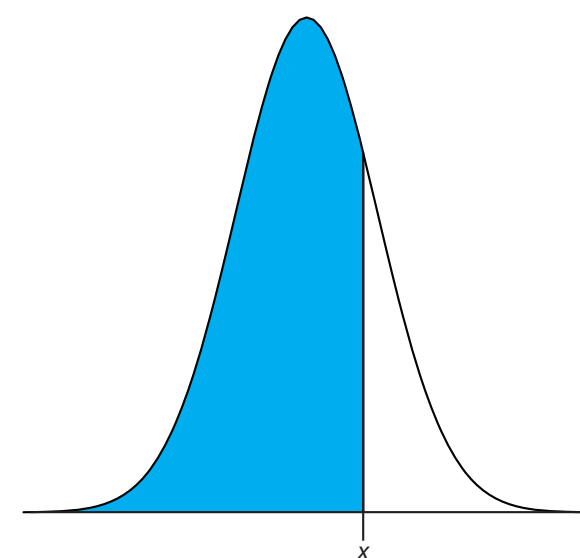
```
> pnorm(243,250,5)
[1] 0.08075666
> pnorm(262,250,5)
[1] 0.9918025
```

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.

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

where  $p$  is the probability of  $X$  occurring in one trial. Using [Example 4.10](#) to test this function, we find that the probability of finding no atoms of  $^{13}\text{C}$  atoms in a molecule of cholesterol,  $\text{C}_{27}\text{H}_{44}\text{O}$  is



**Figure 4.24** Shown in blue is the area returned by the function

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

```
> 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.

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

To find the percentage of cholesterol molecules containing 0, 1, or 2 atoms of  $^{13}\text{C}$ , 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 to help you complete 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
- 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. To do this, 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 must use an  $F$ -test to determine whether the variances for the two data sets are equal or unequal. Our null hypothesis is that the variances are equal,  $s_{\text{Set 1}}^2 = s_{\text{Set 2}}^2$ , and our alternative hypothesis is that the variances are not equal,  $s_{\text{Set 1}}^2 \neq s_{\text{Set 2}}^2$ .

The command for a two-tailed  $F$ -test in R, which is our choice for this problem, is

$$\text{var.test}(X, Y)$$

where  $X$  and  $Y$  are the objects containing the data sets. [Figure 4.25](#) shows the output from an R session to solve this problem.

R does not provide the critical value for  $F(0.05, 6, 4)$ . Instead it reports the 95% confidence interval for  $F_{\text{exp}}$ . 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.

For a one-tailed  $F$ -test the command is one of the following

```
var.test(X, Y, alternative = "greater")
```

```
var.test(X, Y, alternative = "less")
```

where "greater" is used when the alternative hypothesis is  $s_x^2 > s_y^2$ , and "less" is used when the alternative hypothesis is

$$s_x^2 < s_y^2.$$



```
> 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

R calculates  $F_{\text{exp}}$  as  $(s_X)^2/(s_Y)^2$ . If we use the command

```
var.test(penny2, penny1)
```

the output will give R as 0.534 and the 95% confidence interval as 0.0858 to 4.912. Because the expected value for  $F_{\text{exp}}$  of 1.00 falls within the confidence interval, we retain the null hypothesis of equal variances.

**Figure 4.25** 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 the value  $\alpha$  as a p-value*). The 95% confidence interval is the range of values for  $F_{\text{exp}}$  that can be 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.

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. Our null hypothesis is that there is no difference between the means,  $\bar{X}_{\text{Set 1}} = \bar{X}_{\text{Set 2}}$ , and our alternative hypothesis is that there is a difference between the means,  $\bar{X}_{\text{Set 1}} \neq \bar{X}_{\text{Set 2}}$ . In R there is a single command for all two-sample  $t$ -tests. The basic syntax for a two-tailed unpaired  $t$ -test with unequal variances is

```
t.test(X, Y, mu = 0, paired = FALSE, var.equal = FALSE)
```

where  $X$  and  $Y$  are the objects containing 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 a null hypothesis of  $\bar{X}_{\text{Set 1}} = \bar{X}_{\text{Set 2}}$  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 that you intend an unpaired two-tailed  $t$ -test of the null hypothesis that  $X = Y$  with unequal variances. [Figure 4.26](#) shows the output of an R session for this problem.

The p-value of 0.2116 means that there is a 21.16% probability of incorrectly rejecting the null hypothesis. The 95% confidence interval of -0.024 to 0.0958, which is for the difference between the sample means, includes the expected value of zero. Both ways of looking at the results of the  $t$ -test 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.

To complete a one-sided  $t$ -test, include the command

```
alternative = “greater”
```

or

```
alternative = “less”
```

A one-sided paired  $t$ -test that the difference between two samples is greater than 0 becomes

```
t.test(X, Y, paired = TRUE, alternative = “greater”)
```

```

> 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 4.26** 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 the value  $\alpha$  as a  $p$ -value*). The 95% confidence interval is the range of values for the difference between the means that can be 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.

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

```
> 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)
```

where  $X$  is the object containing 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 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. [Figure 4.27](#) 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.

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 4.27** 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.

## VISUALIZING DATA

One of the more useful features of R is the ability to visualize your data. Visualizing your data is important because it provides you with an intuitive feel for your data that can help you 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 your data, however, can help you determine whether your data is normally distributed—a requirement for most of the significance tests in this chapter—and can help you identify potential outliers. There are many useful ways to look at your data, four of which we consider here.

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

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

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 display 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 4.28a](#) shows the result of entering

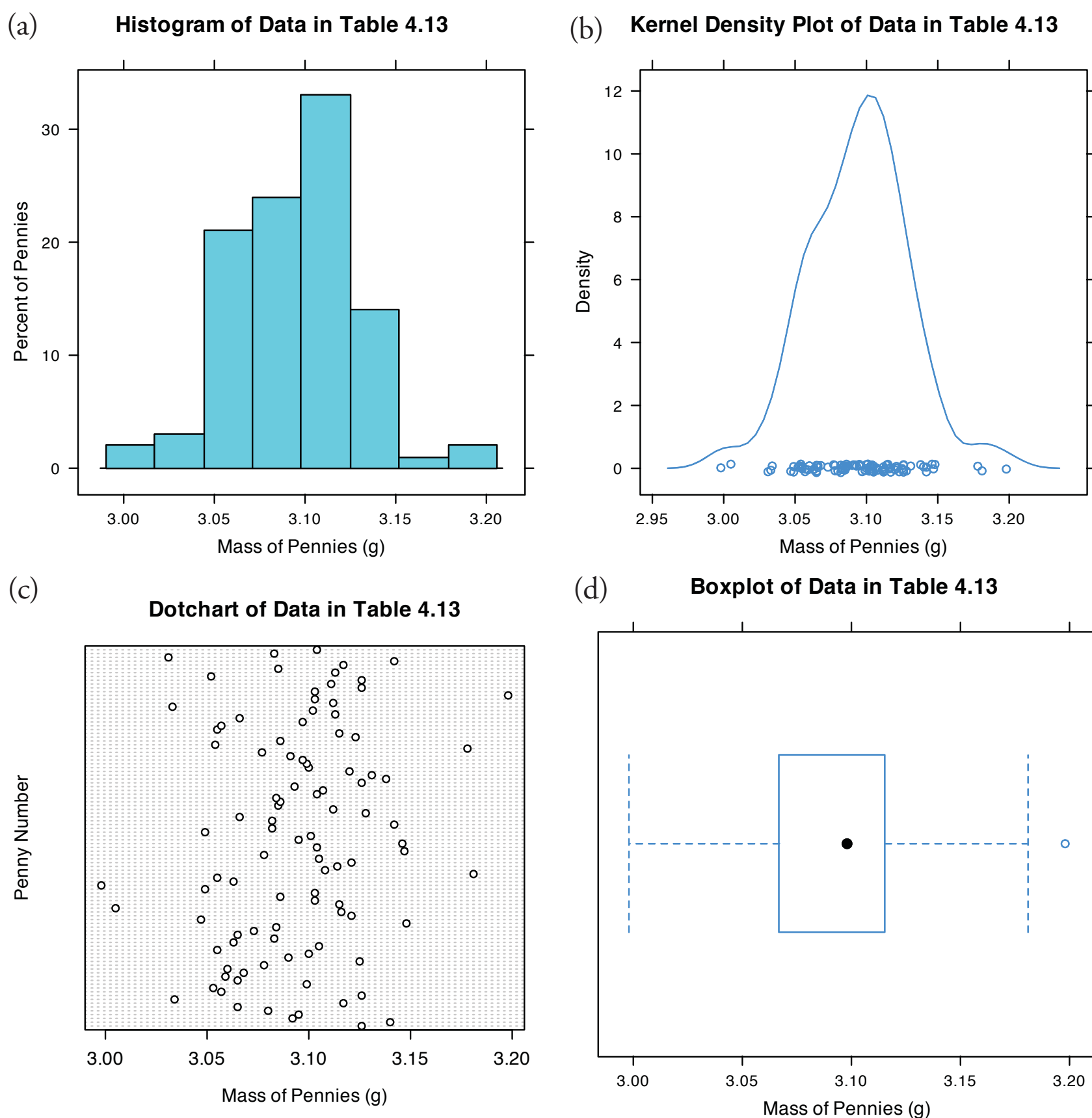
```
> histogram(penny, type = "percent", xlab = "Mass (g)",
  ylab = "Percent of Pennies", main = "Histogram of Data in Table
  4.13")
```

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

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

You can download the file “Penny.Rdata” from the textbook's [web site](#).

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



**Figure 4.28** Four different 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.

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 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 is basically 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 4.28b](#) 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 can 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 must be random. A **DOT CHART** provides a simple visual display that allows us look for non-random trends. [Figure 4.28c](#) 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 of sampling. 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 4.28d](#) shows the result of entering

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

The black dot ( $\bullet$ ) is the data set's median. The rectangular box shows the range of masses for 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  $\pm 1.5 \times \text{IQR}$  of the rectangular box. Potential outliers are shown as open circles ( $\circ$ ). For normally distributed data the median will be near the center of the box and the whiskers will be equidistant from the box. As is often true 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 4.28d](#) is consistent with the histogram (Figure 4.28a) 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

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

$$3.068 - 1.5 \times \text{IQR} = 3.068 - 1.5 \times 0.047 = 2.9975$$

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

$$3.115 + 1.5 \times \text{IQR} = 3.115 + 1.5 \times 0.047 = 3.1855$$

which for this data is 3.181 g.

( $n = 100$ ) 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.

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

## 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 affecting its

precision. A propagation of uncertainty allows us to estimate how these determinate and indeterminate errors will 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 precisions. To compare two sets of data you must first determine whether the data is paired or unpaired. For unpaired data you must also decide if the standard deviations can be pooled. 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 when deciding to reject an outlier.

Finally, the detection limit is a statistical statement about the smallest amount of analyte that 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

- 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.

- A determination of acetaminophen in 10 separate tablets of Excedrin Extra Strength Pain Reliever gives the following results (in mg).<sup>16</sup>

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  $\bar{X}$  and  $s^2$  are good approximations for  $\mu$  and  $\sigma^2$ , and that the population is normally distributed, what per-

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](#)

<sup>16</sup> Simonian, M. H.; Dinh, S.; Fray, L. A. *Spectroscopy* **1993**, 8(6), 37–47.

centage of tablets contain more than the standard amount of 250 mg acetaminophen per tablet?

3. Salem and Galan developed a new method for determining the amount of morphine hydrochloride in tablets.<sup>17</sup> 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	

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](#)

(a) For each dosage, calculate the mean and standard deviation for the mg of morphine hydrochloride per tablet. (b) For each dosage level, assuming that  $\bar{X}$  and  $s^2$  are good approximations for  $\mu$  and  $\sigma^2$ , 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.<sup>18</sup> 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 the spike recoveries fell within the range  $100 \pm 15\%$ . The spike recoveries for one method are shown here.

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 will be within the accepted range?

5. The formula weight (*FW*) of a gas can be determined using the following form of the ideal gas law

$$FW = \frac{gRT}{PV}$$

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.

<sup>17</sup> Salem, I. I.; Galan, A. C. *Anal. Chim. Acta* 1993, 283, 334–337.

<sup>18</sup> Daskalakis, K. D.; O'Connor, T. P.; Crecelius, E. A. *Environ. Sci. Technol.* 1997, 31, 2303–2306.



In a typical analysis the following data are obtained (with estimated uncertainties in parentheses)

$$g = 0.118 \text{ g } (\pm 0.002 \text{ g})$$

$$R = 0.082056 \text{ L atm mol}^{-1} \text{ K}^{-1} (\pm 0.000001 \text{ L atm mol}^{-1} \text{ K}^{-1})$$

$$T = 298.2 \text{ K } (\pm 0.1 \text{ K})$$

$$P = 0.724 \text{ atm } (\pm 0.005 \text{ atm})$$

$$V = 0.250 \text{ L } (\pm 0.005 \text{ L})$$

- (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  $\text{Mn}^{2+}$  a 0.250 g sample of Mn is dissolved in 10 mL of concentrated  $\text{HNO}_3$  (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  $\text{HNO}_3$ , 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 ( $\pm 0.01$  g/mol) was prepared in the following manner. A sample of the compound and its container has a mass of 23.5811 grams. A portion of the compound was transferred to a 100-mL volumetric flask and diluted to volume. The mass of the compound and container after the transfer is 22.1559 grams. Calculate the compound's molarity and estimate its uncertainty by a propagation of uncertainty.
8. Show using a propagation of uncertainty that the standard error of the mean for  $n$  determinations is  $s/\sqrt{n}$ .
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 that we can measure on an analytical balance that has a tolerance of  $\pm 0.1$  mg, if the relative error must be less than 0.1%?

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](#)

11. Which of the following is the best way to dispense 100.0 mL of a reagent: (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  $\bar{X}$  but will have no effect on  $s$ .
14. Obtain a sample of a metal 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 that it displaces. This can be done 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 5 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  $^{13}\text{C}$  atoms per molecule is 1.00? What percentage of such molecules will have no atoms of  $^{13}\text{C}$ ?
16. In Example 4.10 we determined the probability that a molecule of cholesterol,  $\text{C}_{27}\text{H}_{44}\text{O}$ , had no atoms of  $^{13}\text{C}$ . (a) Calculate the probability that a molecule of cholesterol, has 1 atom of  $^{13}\text{C}$ . (b) What is the probability that a molecule of cholesterol will have two or more atoms of  $^{13}\text{C}$ ?
17. Berglund and Wichardt investigated the quantitative determination of Cr in high-alloy steels using a potentiometric titration of Cr(VI)<sup>19</sup>.

<sup>19</sup> Berglund, B.; Wichardt, C. *Anal. Chim. Acta* **1990**, *236*, 399–410.

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](#)

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 for determining trace levels of atmospheric gases.<sup>20</sup> An analysis of a sample containing 40.0 parts per thousand (ppt) 2-chloroethylsulfide yielded 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.<sup>21</sup> 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

Determine whether there is a significant difference between the mean and the expected value at  $\alpha = 0.05$ .

20 Ketkar, S. N.; Dulak, J. G.; Dheandhanou, S.; Fite, W. L. *Anal. Chim. Acta* **1996**, *330*, 267–270.

21 Monna, F.; Mathieu, D.; Marques, A. N.; Lancelot, J.; Bernat, M. *Anal. Chim. Acta* **1996**, *330*, 107–116.

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](#)

21. A 2.6540-g sample of an iron ore, known to be 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.<sup>22</sup> 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. Results obtained using HNO<sub>3</sub> and using a 1 + 3 mixture of HNO<sub>3</sub> and HCl are shown here. All concentrations are given as ng Hg/g sample.

HNO <sub>3</sub>	161	165	160	167	166	
1+3 HNO <sub>3</sub> -HCl	159	145	140	147	143	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<sub>2</sub>. Rayleigh approached this experiment in two ways: first by taking atmospheric air and removing all O<sub>2</sub> and H<sub>2</sub>; and second, by chemically producing N<sub>2</sub> by decomposing nitrogen containing compounds (NO, N<sub>2</sub>O, and NH<sub>4</sub>NO<sub>3</sub>) and again removing all O<sub>2</sub> and H<sub>2</sub>. Following are his results for the density of N<sub>2</sub>, published in *Proc. Roy. Soc.* **1894**, *LV*, 340 (publication 210) (all values are for grams of gas at an equivalent volume, pressure, and temperature).<sup>23</sup>

Atmospheric	2.310 17	2.309 86	2.310 10	2.310 01
Origin:	2.310 24	2.310 10	2.310 28	
Chemical	2.301 43	2.298 90	2.298 16	2.301 82
Origin:	2.298 69	2.299 40	2.298 49	2.298 89

Explain why this data led Rayleigh to look for, and discover Ar.

24. Gács and Ferraroli reported a method for monitoring the concentration of SO<sub>2</sub> in air.<sup>24</sup> 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.

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, I.; Ferraroli, R. *Anal. Chim. Acta* **1992**, *269*, 177–185.

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](#)

Shown here is a summary of their results with  $\text{SO}_2$  concentrations reported in  $\mu\text{L}/\text{m}^3$ .

standard	21.62	22.20	24.27	23.54
method:	24.25	23.09	21.02	
new	21.54	20.51	22.31	21.30
method:	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 absorbencies for a series of standard dichromate solutions obtained from the National Institute of Standards and Technology. Absorbencies 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

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](#)

26. Maskarinec and co-workers investigated the stability of volatile organics in environmental water samples.<sup>25</sup> Of particular interest was establishing proper conditions for maintaining the sample's integrity between its collection and 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 (in days) for the holding time of nine organic compounds in surface water.

	Ascorbic Acid	Sodium Bisulfate
methylene chloride	77	62
carbon disulfide	23	54
trichloroethane	52	51
benzene	62	42
1,1,2-trichloroethane	57	53
1,1,2,2-tetrachlorethane	33	85
tetrachloroethene	41	63

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.

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 Kvalheim reported a new method for determining the weight percent of kalonite in complex clay minerals using X-ray diffraction.<sup>26</sup> To test the method, nine samples containing known amounts of kalonite were prepared and analyzed. The results (as % w/w kalonite) 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.<sup>27</sup> 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 Boehringer 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  $\alpha = 0.05$ .

29. Alexiev and colleagues describe an improved photometric method for determining  $\text{Fe}^{3+}$  based on its ability to catalyze the oxidation of sulphanic acid by  $\text{KIO}_4$ .<sup>28</sup> As part of their study the concentration of  $\text{Fe}^{3+}$

<sup>26</sup> Karstang, T. V.; Kvalheim, O. M. *Anal. Chem.* **1991**, *63*, 767–772.

<sup>27</sup> Mizutani, F.; Yabuki, S.; Asai, M. *Anal. Chim. Acta* **1991**, *245*, 145–150.

<sup>28</sup> Alexiev, A.; Rubino, S.; Deyanova, M.; Stoyanova, A.; Sicilia, D.; Perez Bendito, D. *Anal. Chim. Acta*, **1994**, *295*, 211–219.

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](#)

in human serum samples was determined by the improved method and the standard method. The results, with concentrations in  $\mu\text{mol/L}$ , 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  $\mu\text{g/mL}$ .<sup>29</sup>

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 at a significance level of  $\alpha = 0.05$ . Use all three methods—Dixon's  $Q$ -test, Grubb's test, and Chauvenet's criterion—and compare the results to each other.

31. When copper metal and powdered sulfur are placed in a crucible and ignited, the product is a sulfide with an empirical formula of  $\text{Cu}_x\text{S}$ . The value of  $x$  can be determined by weighing the Cu and S before ignition, and finding the mass of  $\text{Cu}_x\text{S}$  when the reaction is complete (any excess sulfur leaves as  $\text{SO}_2$ ). The following table shows the Cu/S ratios from 62 such experiments.

<sup>29</sup> 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.

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](#)

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 and standard deviation for this data. (b) Construct a histogram for this data. From a visual inspection of your histogram, does the data appear to be 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  $\pm 1s$ ? Does this support your answer to the previous question? (d) Assuming that  $\bar{X}$  and  $s^2$  are good approximations for  $\mu$  and  $\sigma^2$ , what percentage of all experimentally determined Cu/S ratios will 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  $\alpha = 0.01$ .

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,  $C_t$ , 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.<sup>30</sup> Each sample was analyzed 18 times.

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

<sup>30</sup> Burns, M. J.; Nixon, G. J.; Foy, C. A.; Harris, N. *BMC Biotechnol.* **2005**, *5*:31 ([open access publication](#)).

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](#)



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 statistically 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.

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](#)

## 4L Solutions to Practice Exercises

### Practice Exercise 4.1

*Mean:* To find the mean we sum up 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  $1.40 \times 10^2$  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 141 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  $\text{Cu}^{2+}$  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  $\text{HNO}_3$  used to dissolve the copper does not factor into our calculation. The concentration of  $\text{Cu}^{2+}$  is

$$\frac{0.9775 \text{ g Cu}}{0.5000 \text{ L}} \times \frac{1.000 \text{ mL}}{250.0 \text{ mL}} \times \frac{10^3 \text{ mg}}{\text{g}} = 7.820 \text{ mg Cu}^{2+} / \text{L}$$

Having found the concentration of  $\text{Cu}^{2+}$  we continue on to complete 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  $\text{Cu}^{2+}$  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  $\text{Cu}^{2+}$  is  $7.820 \text{ mg/L} \pm 0.047 \text{ mg/L}$ .

Click [here](#) to return to the chapter.

### Practice Exercise 4.3

The first step is to calculate the absorbance, which is

$$A = -\log \frac{P}{P_0} = -\log \frac{1.50 \times 10^2}{3.80 \times 10^2} = -\log(0.3947) = 0.4037 \approx 0.404$$

Having found the absorbance we continue on to complete the propagation of uncertainty. First, we find the uncertainty for the ratio  $P/P_0$ .

$$\frac{u_{P/P_0}}{P/P_0} = \sqrt{\left(\frac{15}{3.80 \times 10^2}\right)^2 + \left(\frac{15}{1.50 \times 10^2}\right)^2} = 1.075 \times 10^{-2}$$

Finally, from [Table 4.10](#) the uncertainty in the absorbance is

$$u_A = 0.4343 \times \frac{u_{P/P_0}}{P/P_0} = (0.4343) \times (1.075 \times 10^{-2}) = 4.669 \times 10^{-3}$$

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

$$6.4 \times 10^{-5} = \left(\frac{0.028}{23.41}\right)^2 + \left(\frac{u_{k_A}}{0.186}\right)^2$$

$$6.257 \times 10^{-5} = \left(\frac{u_{k_A}}{0.186}\right)^2$$

Solving for  $u_{k_A}$  gives its value as  $1.47 \times 10^{-3}$ , or  $\pm 0.0015 \text{ ppm}^{-1}$ .

Click [here](#) to return to the chapter.

---

#### Practice Exercise 4.5

To find the percentage of tablets containing less than 245 mg of aspirin we calculate the deviation,  $z$ ,

$$z = \frac{245 - 250}{5} = -1.00$$

and look up the corresponding probability in Appendix 3A, obtaining a value of 15.87%. To find the percentage of tablets containing 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 aspirin 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. The confidence interval is

$$249.9 \pm \frac{1.96 \times 5}{\sqrt{7}} = 249.9 \pm 3.7 \approx 250 \text{ mg} \pm 4 \text{ mg}$$

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 using the values for  $t(0.05, 60)$  and  $t(0.05, 100)$  and 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  $3.117 \text{ g} \pm 0.047 \text{ g}$  and  $3.081 \text{ g} \pm 0.046 \text{ g}$ . 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  $H_0: \bar{X} = \mu$  and the alternative hypothesis is  $H_A: \bar{X} \neq \mu$ . The mean and standard deviation for the data are 99.26% and 2.35%, respectively. The value for  $t_{\text{exp}}$  is

$$t_{\text{exp}} = \frac{|100.0 - 99.26| \sqrt{7}}{2.35} = 0.833$$

and the critical value for  $t(0.05, 6)$  is 0.836. Because  $t_{\text{exp}}$  is less than  $t(0.05, 6)$  we retain the null hypothesis and have no evidence for a significant difference between  $\bar{X}$  and  $\mu$ .

Click [here](#) to return to the chapter.

---

**Practice Exercise 4.9**

The standard deviations for Lot 1 is 6.451 mg, and 7.849 mg for Lot 2. The null and alternative hypotheses are

$$H_0 : s_{\text{Lot 1}}^2 = s_{\text{Lot 2}}^2 \quad H_A : s_{\text{Lot 1}}^2 \neq s_{\text{Lot 2}}^2$$

and the value of  $F_{\text{exp}}$  is

$$F_{\text{exp}} = \frac{(7.849)^2}{(6.451)^2} = 1.480$$

The critical value for  $F(0.05, 5, 6)$  is 5.988. Because  $F_{\text{exp}} < F(0.05, 5, 6)$ , 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 will use an unpaired t-test of the null hypothesis  $H_0 : \bar{X}_{\text{Lot 1}} = \bar{X}_{\text{Lot 2}}$  and the alternative hypothesis  $H_A : \bar{X}_{\text{Lot 1}} \neq \bar{X}_{\text{Lot 2}}$ . Because there is no evidence suggesting a difference in the variances (see [Practice Exercise 4.9](#)) we pool the standard deviations, obtaining an  $s_{\text{pool}}$  of

$$s_{\text{pool}} = \sqrt{\frac{(7-1)(6.451)^2 + (6-1)(7.849)^2}{7+6-2}} = 7.121$$

The means for the two samples are 249.57 mg for Lot 1 and 249.00 mg for Lot 2. The value for  $t_{\text{exp}}$  is

$$t_{\text{exp}} = \frac{|249.57 - 249.00|}{7.121} \times \sqrt{\frac{7 \times 6}{7+6}} = 0.1439$$

The critical value for  $t(0.05, 11)$  is 2.204. Because  $t_{\text{exp}}$  is less than  $t(0.05, 11)$ , we retain the null hypothesis and find no evidence at  $\alpha = 0.05$  for 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 standard deviation for the concentration of  $\text{Zn}^{2+}$  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. An  $F$ -test of the variances gives an  $F_{\text{exp}}$  of 1.493 and an  $F(0.05, 5, 5)$  of 7.146. Because  $F_{\text{exp}}$  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  $s_{\text{pool}}$  of 0.1582

mg/L. An unpaired  $t$ -test gives  $t_{\text{exp}}$  as 0.8025. Because  $t_{\text{exp}}$  is smaller than  $t(0.05, 11)$ , which is 2.204, we have no evidence that there is a difference in the concentration of  $\text{Zn}^{2+}$  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  $\text{Zn}^{2+}$  at the air-water interface and at the sediment-water interface for each location.

$$d_i = \left( [\text{Zn}^{2+}]_{\text{air-water}} \right)_i - \left( [\text{Zn}^{2+}]_{\text{sed-water}} \right)_i$$

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 alternative hypothesis are

$$H_0 : \bar{d} = 0 \quad H_A : \bar{d} \neq 0$$

and the value of  $t_{\text{exp}}$  is

$$t_{\text{exp}} = \frac{|0.07333| \sqrt{6}}{0.04410} = 4.073$$

Because  $t_{\text{exp}}$  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  $\text{Zn}^{2+}$  between the air-water interface and the sediment-water interface.

The difference in the concentration of  $\text{Zn}^{2+}$  between locations is much larger than the difference in the concentration of  $\text{Zn}^{2+}$  between the interfaces. Because our interest is in studying differences 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 [here](#) to return to the chapter.

---

### Practice Exercise 4.12

You will find small differences between the values given here for  $t_{\text{exp}}$  and  $F_{\text{exp}}$ , and for those values shown with the worked solutions in the chapter. These differences arise because Excel does not round off the results of intermediate calculations.

The two snapshots of Excel spreadsheets shown in [Figure 4.29](#) provide solutions to these two examples.

Click [here](#) to return to the chapter.

---

Example 4.20	Analyst A	Analyst B	Example 4.21	Microbiological	Electrochemical
	86.82	81.01		129.5	132.3
	87.04	86.15		89.6	91.0
	86.93	81.73		76.6	73.6
	87.01	83.19		52.2	58.2
	86.20	80.27		110.8	104.2
	87.00	83.94		50.4	49.9
				72.4	82.1
				141.4	154.1
				75.0	73.4
				34.1	38.1
				60.3	60.1

F-Test Two-Sample for Variances			
	Analyst B	Analyst A	
Mean	82.715	86.8333333	
Variance	4.67615	0.10246667	
Observations	6	6	
df	5	5	
F	45.6358165		
P(F<=f) one-tail	0.0003574		
F Critical one-tail	5.05032906		

t-Test: Two-Sample Assuming Unequal Variances			
	Analyst A	Analyst B	
Mean	86.8333333	82.715	
Variance	0.10246667	4.67615	
Observations	6	6	
Hypothesized Mean Difference	0		
df	5		
t Stat	4.6147271		
P(T<=t) one-tail	0.00288143		
t Critical one-tail	2.01504837		
P(T<=t) two-tail	0.00576286		
t Critical two-tail	2.57058183		

t-Test: Paired Two Sample for Means			
	Microbiological	Electrochemical	
Mean	81.11818182	83.36363636	
Variance	1148.547636	1243.008545	
Observations	11	11	
Pearson Correlation	0.987510302		
Hypothesized Mean Difference	0		
df	10		
t Stat	-1.322463865		
P(T<=t) one-tail	0.107730346		
t Critical one-tail	1.812461102		
P(T<=t) two-tail	0.215460691		
t Critical two-tail	2.228138842		

Figure 4.29 Excel's output for the data in Practice Exercise 4.12.

### 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  $t_{\text{exp}}$  and  $F_{\text{exp}}$ , and for 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

```

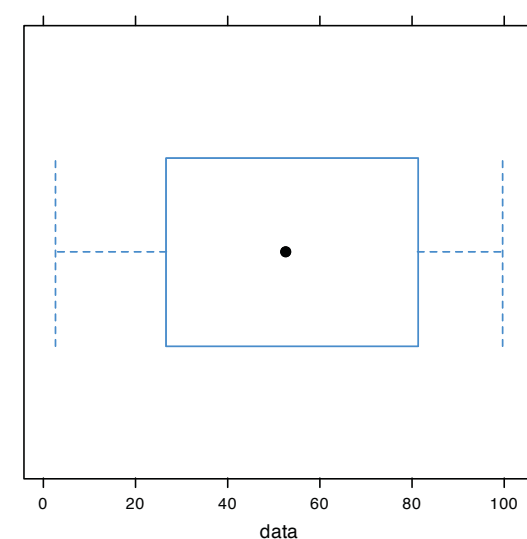
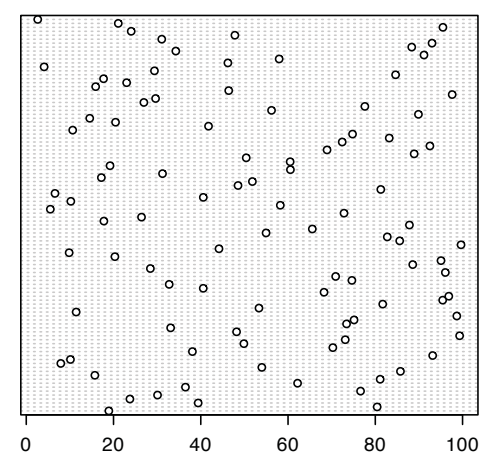
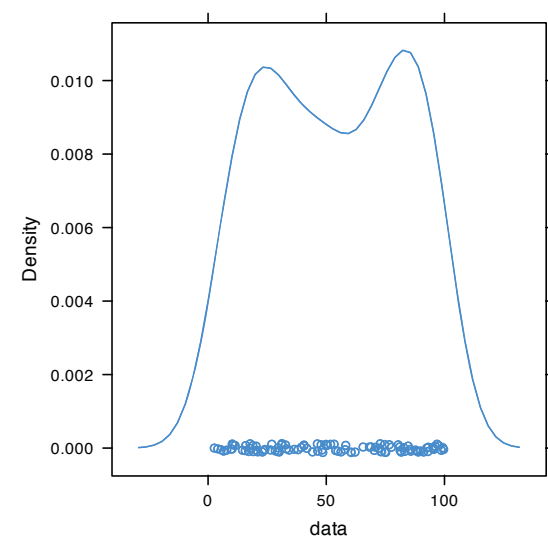
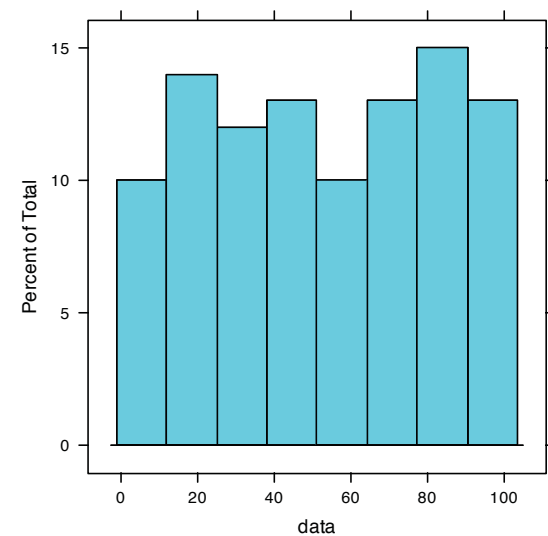


```
[31] 68.251721 40.567993 32.761695 74.635385 70.914957
[36] 96.054750 28.448719 88.580214 95.059215 20.316015
[41]  9.828515 44.172774 99.648405 85.593858 82.745774
[46] 54.963426 65.563743 87.820985 17.791443 26.417481
[51] 72.832037  5.518637 58.231329 10.213343 40.581266
[56]  6.584000 81.261052 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 4.30 shows the four plots. The histogram divides the data into eight bins, each containing 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, indicating that the individual members of this sample were drawn randomly 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 4.30** Plots generated using R to solve Practice Exercise 4.13.