Chemistry

Modern Analytical Chemistry

David Harvey
DePauw University

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Lisbon London Madrid
Mexico City Milan New Delhi Seoul Singapore Sydney Taipei Toronto

McGraw-Hill Higher Education
*A Division of The **McGraw-Hill** Companies*

MODERN ANALYTICAL CHEMISTRY

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 KGP/KGP 0 9 8 7 6 5 4 3 2 1 0

ISBN 0–07–237547–7

Vice president and editorial director: *Kevin T. Kane*
Publisher: *James M. Smith*
Sponsoring editor: *Kent A. Peterson*
Editorial assistant: *Jennifer L. Bensink*
Developmental editor: *Shirley R. Oberbroeckling*
Senior marketing manager: *Martin J. Lange*
Senior project manager: *Jayne Klein*
Production supervisor: *Laura Fuller*
Coordinator of freelance design: *Michelle D. Whitaker*
Senior photo research coordinator: *Lori Hancock*
Senior supplement coordinator: *Audrey A. Reiter*
Compositor: *Shepherd, Inc.*
Typeface: *10/12 Minion*
Printer: *Quebecor Printing Book Group/Kingsport*

Freelance cover/interior designer: *Elise Lansdon*
Cover image: *© George Diebold/The Stock Market*
Photo research: *Roberta Spieckerman Associates*

Colorplates: Colorplates 1–6, 8, 10: © David Harvey/Marilyn E. Culler, photographer; Colorplate 7: Richard Megna/Fundamental Photographs; Colorplate 9: © Alfred Pasieka/Science Photo Library/Photo Researchers, Inc.; Colorplate 11: From H. Black, *Environ. Sci. Technol.,* **1996,** *30,* 124A. Photos courtesy D. Pesiri and W. Tumas, Los Alamos National Laboratory; Colorplate 12: Courtesy of Hewlett-Packard Company; Colorplate 13: © David Harvey.

Library of Congress Cataloging-in-Publication Data

Harvey, David, 1956–
Modern analytical chemistry / David Harvey. — 1st ed.
p. cm.
Includes bibliographical references and index.
ISBN 0–07–237547–7
1. Chemistry, Analytic. I. Title.
QD75.2.H374 2000
543—dc21 99–15120
CIP

INTERNATIONAL EDITION ISBN 0–07–116953–9

www.mhhe.com

Contents

Chapter 6
Equilibrium Chemistry 135

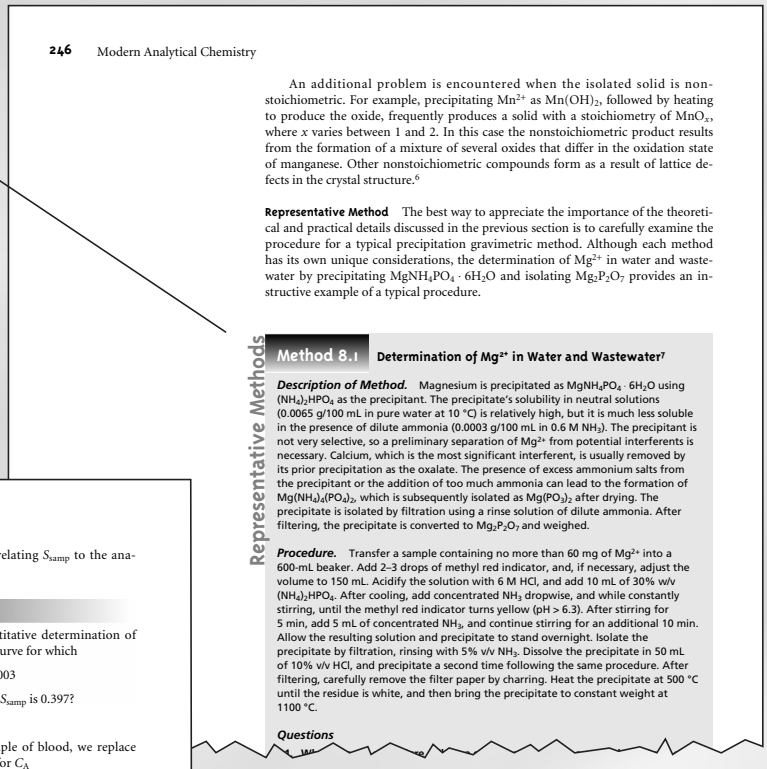
Chapter 7
Obtaining and Preparing Samples for Analysis 179

A Guide to Using This Text

. . . in Chapter

Representative Methods

Annotated methods of typical analytical procedures link theory with practice. The format encourages students to think about the design of the procedure and why it works.



246 Modern Analytical Chemistry

An additional problem is encountered when the isolated solid is nonstoichiometric. For example, precipitating Mn^{2+} as $Mn(OH)_2$, followed by heating to produce the oxide, frequently produces a solid with a stoichiometry of MnO_x, where x varies between 1 and 2. In this case the nonstoichiometric product results from the formation of a mixture of several oxides that differ in the oxidation state of manganese. Other nonstoichiometric compounds form as a result of lattice defects in the crystal structure.[6]

Representative Method The best way to appreciate the importance of the theoretical and practical details discussed in the previous section is to carefully examine the procedure for a typical precipitation gravimetric method. Although each method has its own unique considerations, the determination of Mg^{2+} in water and wastewater by precipitating $MgNH_4PO_4 \cdot 6H_2O$ and isolating $Mg_2P_2O_7$ provides an instructive example of a typical procedure.

Representative Methods

Method 8.1 **Determination of Mg^{2+} in Water and Wastewater[7]**

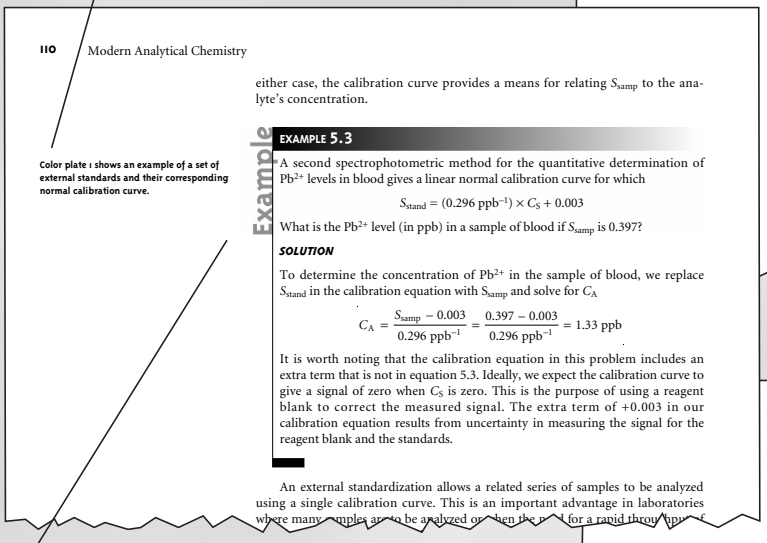
Description of Method. Magnesium is precipitated as $MgNH_4PO_4 \cdot 6H_2O$ using $(NH_4)_2HPO_4$ as the precipitant. The precipitate's solubility in neutral solutions (0.0065 g/100 mL in pure water at 10 °C) is relatively high, but it is much less soluble in the presence of dilute ammonia (0.0003 g/100 mL in 0.6 M NH_3). The precipitant is not very selective, so a preliminary separation of Mg^{2+} from potential interferents is necessary. Calcium, which is the most significant interferent, is usually removed by its prior precipitation as the oxalate. The presence of excess ammonium salts from the precipitant or the addition of too much ammonia can lead to the formation of $Mg(NH_4)_4(PO_4)_2$, which is subsequently isolated as $Mg(PO_3)_2$ after drying. The precipitate is isolated by filtration using a rinse solution of dilute ammonia. After filtering, the precipitate is converted to $Mg_2P_2O_7$ and weighed.

Procedure. Transfer a sample containing no more than 60 mg of Mg^{2+} into a 600-mL beaker. Add 2–3 drops of methyl red indicator, and, if necessary, adjust the volume to 150 mL. Acidify the solution with 6 M HCl, and add 10 mL of 30% w/v $(NH_4)_2HPO_4$. After cooling, add concentrated NH_3 dropwise, and while constantly stirring, until the methyl red indicator turns yellow (pH > 6.3). After stirring for 5 min, add 5 mL of concentrated NH_3, and continue stirring for an additional 10 min. Allow the resulting solution and precipitate to stand overnight. Isolate the precipitate by filtration, rinsing with 5% v/v NH_3. Dissolve the precipitate in 50 mL of 10% v/v HCl, and precipitate a second time following the same procedure. After filtering, carefully remove the filter paper by charring. Heat the precipitate at 500 °C until the residue is white, and then bring the precipitate to constant weight at 1100 °C.

Questions

Margin Notes

Margin notes direct students to colorplates located toward the middle of the book



110 Modern Analytical Chemistry

either case, the calibration curve provides a means for relating S_{samp} to the analyte's concentration.

Example

EXAMPLE 5.3

A second spectrophotometric method for the quantitative determination of Pb^{2+} levels in blood gives a linear normal calibration curve for which

$$S_{stand} = (0.296\ ppb^{-1}) \times C_S + 0.003$$

What is the Pb^{2+} level (in ppb) in a sample of blood if S_{samp} is 0.397?

SOLUTION

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

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

It is worth noting that the calibration equation in this problem includes an extra term that is not in equation 5.3. Ideally, we expect the calibration curve to give a signal of zero when C_S is zero. This is the purpose of using a reagent blank to correct the measured signal. The extra term of +0.003 in our calibration equation results from uncertainty in measuring the signal for the reagent blank and the standards.

An external standardization allows a related series of samples to be analyzed using a single calibration curve. This is an important advantage in laboratories where many samples are to be analyzed or when the need for a rapid throughput of

Color plate 1 shows an example of a set of external standards and their corresponding normal calibration curve.

There is a serious limitation, however, to an external standardization. The relationship between S_{stand} and C_S in equation 5.3 is determined when the analyte is present in the external standard's matrix. In using an external standardization, we assume that any difference between the matrix of the standards and the sample's matrix has no effect on the value of k. A proportional determinate error is introduced when differences between the two matrices cannot be ignored. This is shown in Figure 5.4, where the relationship between the signal and the amount of analyte is shown for both the sample's matrix and the standard's matrix. In this example, using a normal calibration curve results in a negative determinate error. When matrix problems are expected, an effort is made to match the matrix of the standards to that of the sample. This is known as **matrix matching.** When the sample's matrix is unknown, the matrix effect must be shown to be negligible, or an alternative method of standardization must be used. Both approaches are discussed in the following sections.

matrix matching
Adjusting the matrix of an external standard so that it is the same as the matrix of the samples to be analyzed.

method of standard additions
A standardization in which aliquots of a standard solution are added to the sample.

5B.4 Standard Additions

The complication of matching the matrix of the standards to that of the sample can be avoided by conducting the standardization in the sample. This is known as the **method of standard additions.** The simplest version of a standard addition is shown in Figure 5.5. A volume, V_o, of sample is diluted to a final volume, V_f, and the signal, S_{samp} is measured. A second identical **aliquot** of sample is

Examples of Typical Problems

Each example problem includes a detailed solution that helps students in applying the chapter's material to practical problems.

Bold-faced Key Terms with Margin Definitions

Key words appear in boldface when they are introduced within the text. The term and its definition appear in the margin for quick review by the student. All key words are also defined in the glossary.

. . . End of Chapter

5E KEY TERMS

aliquot (*p. 111*)
external standard (*p. 109*)
internal standard (*p. 116*)
linear regression (*p. 118*)
matrix matching (*p. 110*)
method of standard additions (*p. 110*)
multiple-point standardization (*p. 109*)
normal calibration curve (*p. 109*)
primary reagent (*p. 106*)
reagent grade (*p. 107*)
residual error (*p. 118*)
secondary reagent (*p. 107*)
single-point standardization (*p. 108*)
standard deviation about the regression (*p. 121*)
total Youden blank (*p. 129*)

5F SUMMARY

In a quantitative analysis, we measure a signal and calculate the amount of analyte using one of the following equations.

$$S_{meas} = kn_A + S_{reag}$$

$$S_{meas} = kC_A + S_{reag}$$

To obtain accurate results we must eliminate determinate errors affecting the measured signal, S_{meas}, the method's sensitivity, k, and any signal due to the reagents, S_{reag}.

To ensure that S_{meas} is determined accurately, we calibrate the equipment or instrument used to obtain the signal. Balances are calibrated using standard weights. When necessary, we can also correct for the buoyancy of air. Volumetric glassware can be calibrated by measuring the mass of water contained or delivered and using the density of water to calculate the true volume. Most instruments have calibration standards suggested by the manufacturer.

An analytical method is standardized by determining its sensitivity. There are several approaches to standardization, including the use of external standards, the method of standard addition, and the use of an internal standard. The most desirable standardization strategy is an external standardization. The method of standard additions, in which known amounts of analyte are added to the sample, is used when the sample's matrix complicates the analysis. An internal standard, which is a species (not analyte) added to all samples and standards, is used when the procedure does not allow for the reproducible handling of samples and standards.

Standardizations using a single standard are common, but also are subject to greater uncertainty. Whenever possible, a multiple-point standardization is preferred. The results of a multiple-point standardization are graphed as a calibration curve. A linear regression analysis can provide an equation for the standardization.

A reagent blank corrects the measured signal for signals due to reagents other than the sample that are used in an analysis. The most common reagent blank is prepared by omitting the sample. When a simple reagent blank does not compensate for all constant sources of determinate error, other types of blanks, such as the total Youden blank, can be used.

5G *Suggested* EXPERIMENTS

Experiments

The following exercises and experiments help connect the material in this chapter to the analytical laboratory.

Calibration—Volumetric glassware (burets, pipets, and volumetric flasks) can be calibrated in the manner described in Example 5.1. Most instruments have a calibration sample that can be prepared to verify the instrument's accuracy and precision. For example, as described in this chapter, a solution of 60.06 ppm $K_2Cr_2O_7$ in 0.0050 M H_2SO_4 should give an absorbance of 0.640 ± 0.010 at a wavelength of 350.0 nm when using 0.0050 M H_2SO_4 as a reagent blank. These exercises also provide practice with using volumetric glassware, weighing samples, and preparing solutions.

Standardization—External standards, standard additions, and internal standards are a common feature of many quantitative analyses. Suggested experiments using these standardization methods are found in later chapters. A good project experiment for introducing external standardization, standard additions, and the importance of the sample's matrix is to explore the effect of pH [illegible] analysis of an acid–base indicator. U[illegible] as an example, external standards ca[illegible] buffer and used to analyze samples [illegible] in the range of 6–10. Results can be [illegible] obtained using a standard addition.

5H REFERENCES

1. Ravey, M. *Spectroscopy* **1990,** *5(7),* 11.
2. de Haseth, J. *Spectroscopy* **1990,** *5(7),* 11.
3. Fresenius, C. R. *A System of Instruction in Quantitative Chemical Analysis.* John Wiley and Sons: New York, 1881.
4. Hillebrand, W. F.; Lundell, G. E. F. *Applied Inorganic Analysis,* John Wiley and Sons: New York, 1953.
5. Van Loon, J. C. *Analytical Atomic Absorption Spectroscopy.* Academic Press: New York, 1980.
6. Murray, R. W. *Anal. Chem.* **1991,** *63,* 271A.
7. For several different viewpoints see (a) Beilby, A. L. *J. Chem. Educ.* **1970,** *47,* 237–238; (b) Lucchesi, C. A. *Am. Lab.* **1980,** *October,* 113–119; (c) Atkinson, G. F. *J. Chem. Educ.* **1982,** *59,* 201–202; (d) Pardue, H. L.; Woo, J. *J. Chem. Educ.* **1984,** *61,* 409–412; (e) Guarnieri, M. *J. Chem. Educ.* **1988,** *65,* 201–203; (f) de Haseth, J. *Spectroscopy* **1990,** *5,* 20–21; (g) Strobel, H. A. *Am. Lab.* **1990,** *October,* 17–24.
8. Hieftje, G. M. *Am. Lab.* **1993,** *October,* 53–61.
9. See, for example, the following laboratory texts: (a) Sorum, C. H.; Lagowski, J. J. *Introduction to Semimicro Qualitative Analysis,* 5th ed. Prentice-Hall: Englewood Cliffs, NJ, 1977.; (b) Shriner, R. L.; Fuson, R. C.; Curtin, D. Y. *The Systematic Identification of Organic Compounds,* 5th ed. John Wiley and Sons: New York, 1964.

5I PROBLEMS

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

   After dissolving the sample in a beaker, remove any solid impurities by passing the solution containing the analyte through filter paper, collecting the solution in a clean Erlenmeyer flask. Rinse the beaker with several small portions of solvent, passing these rinsings through the filter paper, and collecting them in the same Erlenmeyer flask. Finally, rinse the filter paper with several portions of solvent, collecting the rinsings in the same Erlenmeyer flask.

   For a typical concentration method, however, the procedure [illegible]

4. A sample was analyzed to determine the concentration of an analyte. Under the conditions of the analysis, the sensitivity is 17.2 ppm^{-1}. What is the analyte's concentration if S_{meas} is 35.2 and S_{reag} is 0.6?
5. A method for the analysis of Ca^{2+} in water suffers from an interference in the presence of Zn^{2+}. When the concentration of Ca^{2+} is 50 times greater than that of Zn^{2+}, an analysis for Ca^{2+} gives a relative error of –2.0%. What is the value of the selectivity coefficient for this method?
6. The quantitative analysis for reduced glutathione in blood is complicated by the presence of many potential interferents. In one study, when analyzing a solution of 10-ppb glutathione and 1.5-ppb ascorbic acid, the signal was 5.43 times greater than that obtained for the analysis of 10-ppb glutathione.[12] What is the selectivity coefficient for this analysis? The same study found that when analyzing a solution of 350-ppb methionine and 10-ppb glutathione, the

List of Key Terms

The key terms introduced within the chapter are listed at the end of each chapter. Page references direct the student to the definitions in the text.

Summary

The summary provides the student with a brief review of the important concepts within the chapter.

Suggested Experiments

An annotated list of representative experiments is provided from the *Journal of Chemical Education.*

Suggested Readings

Suggested readings give the student access to more comprehensive discussion of the topics introduced within the chapter.

References

The references cited in the chapter are provided so the student can access them for further information.

Problems

A variety of problems, many based on data from the analytical literature, provide the student with practical examples of current research.

Preface

As currently taught, the introductory course in analytical chemistry emphasizes quantitative (and sometimes qualitative) methods of analysis coupled with a heavy dose of equilibrium chemistry. Analytical chemistry, however, is more than equilibrium chemistry and a collection of analytical methods; it is an approach to solving chemical problems. Although discussing different methods is important, that discussion should not come at the expense of other equally important topics. The introductory analytical course is the ideal place in the chemistry curriculum to explore topics such as experimental design, sampling, calibration strategies, standardization, optimization, statistics, and the validation of experimental results. These topics are important in developing good experimental protocols, and in interpreting experimental results. If chemistry is truly an experimental science, then it is essential that all chemistry students understand how these topics relate to the experiments they conduct in other chemistry courses.

Currently available textbooks do a good job of covering the diverse range of wet and instrumental analysis techniques available to chemists. Although there is some disagreement about the proper balance between wet analytical techniques, such as gravimetry and titrimetry, and instrumental analysis techniques, such as spectrophotometry, all currently available textbooks cover a reasonable variety of techniques. These textbooks, however, neglect, or give only brief consideration to, obtaining representative samples, handling interferents, optimizing methods, analyzing data, validating data, and ensuring that data are collected under a state of statistical control.

In preparing this textbook, I have tried to find a more appropriate balance between theory and practice, between "classical" and "modern" methods of analysis, between analyzing samples and collecting and preparing samples for analysis, and between analytical methods and data analysis. Clearly, the amount of material in this textbook exceeds what can be covered in a single semester; it's my hope, however, that the diversity of topics will meet the needs of different instructors, while, perhaps, suggesting some new topics to cover.

The anticipated audience for this textbook includes students majoring in chemistry, and students majoring in other science disciplines (biology, biochemistry, environmental science, engineering, and geology, to name a few), interested in obtaining a stronger background in chemical analysis. It is particularly appropriate for chemistry majors who are not planning to attend graduate school, and who often do not enroll in those advanced courses in analytical chemistry that require physical chemistry as a pre-requisite. Prior coursework of a year of general chemistry is assumed. Competence in algebra is essential; calculus is used on occasion, however, its presence is not essential to the material's treatment.

Key Features of This Textbook

Key features set this textbook apart from others currently available.

- *A stronger emphasis on the evaluation of data.* Methods for characterizing chemical measurements, results, and errors (including the propagation of errors) are included. Both the binomial distribution and normal distribution are presented, and the idea of a confidence interval is developed. Statistical methods for evaluating data include the *t*-test (both for paired and unpaired data), the *F*-test, and the treatment of outliers. Detection limits also are discussed from a statistical perspective. Other statistical methods, such as ANOVA and ruggedness testing, are presented in later chapters.
- *Standardizations and calibrations are treated in a single chapter.* Selecting the most appropriate calibration method is important and, for this reason, the methods of external standards, standard additions, and internal standards are gathered together in a single chapter. A discussion of curve-fitting, including the statistical basis for linear regression (with and without weighting) also is included in this chapter.
- *More attention to selecting and obtaining a representative sample.* The design of a statistically based sampling plan and its implementation are discussed earlier, and in more detail than in other textbooks. Topics that are covered include how to obtain a representative sample, how much sample to collect, how many samples to collect, how to minimize the overall variance for an analytical method, tools for collecting samples, and sample preservation.
- *The importance of minimizing interferents is emphasized.* Commonly used methods for separating interferents from analytes, such as distillation, masking, and solvent extraction, are gathered together in a single chapter.
- *Balanced coverage of analytical techniques.* The six areas of analytical techniques—gravimetry, titrimetry, spectroscopy, electrochemistry, chromatography, and kinetics—receive roughly equivalent coverage, meeting the needs of instructors wishing to emphasize wet methods and those emphasizing instrumental methods. Related methods are gathered together in a single chapter encouraging students to see the similarities between methods, rather than focusing on their differences.
- *An emphasis on practical applications.* Throughout the text applications from organic chemistry, inorganic chemistry, environmental chemistry, clinical chemistry, and biochemistry are used in worked examples, representative methods, and end-of-chapter problems.
- *Representative methods link theory with practice.* An important feature of this text is the presentation of representative methods. These boxed features present typical analytical procedures in a format that encourages students to think about why the procedure is designed as it is.
- *Separate chapters on developing a standard method and quality assurance.* Two chapters provide coverage of methods used in developing a standard method of analysis, and quality assurance. The chapter on developing a standard method includes topics such as optimizing experimental conditions using response surfaces, verifying the method through the blind analysis of standard samples and ruggedness testing, and collaborative testing using Youden's two-sample approach and ANOVA. The chapter on quality assurance covers quality control and internal and external techniques for quality assessment, including the use of duplicate samples, blanks, spike recoveries, and control charts.

- *Problems adapted from the literature.* Many of the in-chapter examples and end-of-chapter problems are based on data from the analytical literature, providing students with practical examples of current research in analytical chemistry.
- *An emphasis on critical thinking.* Critical thinking is encouraged through problems in which students are asked to explain why certain steps in an analytical procedure are included, or to determine the effect of an experimental error on the results of an analysis.
- *Suggested experiments from the Journal of Chemical Education.* Rather than including a short collection of experiments emphasizing the analysis of standard unknowns, an annotated list of representative experiments from the *Journal of Chemical Education* is included at the conclusion of most chapters. These experiments may serve as stand alone experiments, or as starting points for individual or group projects.

The Role of Equilibrium Chemistry in Analytical Chemistry

Equilibrium chemistry often receives a significant emphasis in the introductory analytical chemistry course. While an important topic, its overemphasis can cause students to confuse analytical chemistry with equilibrium chemistry. Although attention to solving equilibrium problems is important, it is equally important for students to recognize when such calculations are impractical, or when a simpler, more qualitative approach is all that is needed. For example, in discussing the gravimetric analysis of Ag^+ as AgCl, there is little point in calculating the equilibrium solubility of AgCl since the concentration of Cl^- at equilibrium is rarely known. It is important, however, to qualitatively understand that a large excess of Cl^- increases the solubility of AgCl due to the formation of soluble silver-chloro complexes. Balancing the presentation of a rigorous approach to solving equilibrium problems, this text also introduces the use of ladder diagrams as a means for providing a qualitative picture of a system at equilibrium. Students are encouraged to use the approach best suited to the problem at hand.

Computer Software

Many of the topics covered in analytical chemistry benefit from the availability of appropriate computer software. In preparing this text, however, I made a conscious decision to avoid a presentation tied to a single computer platform or software package. Students and faculty are increasingly experienced in the use of computers, spreadsheets, and data analysis software; their use is, I think, best left to the personal choice of each student and instructor.

Organization

The textbook's organization can be divided into four parts. Chapters 1–3 serve as an introduction, providing an overview of analytical chemistry (Chapter 1); a review of the basic tools of analytical chemistry, including significant figures, units, and stoichiometry (Chapter 2); and an introduction to the terminology used by analytical chemists (Chapter 3). Familiarity with the material in these chapters is assumed throughout the remainder of the text.

Chapters 4–7 cover a number of topics that are important in understanding how a particular analytical method works. Later chapters are mostly independent of the material in these chapters. Instructors may pick and choose from among the topics

of these chapters, as needed, to support individual course goals. The statistical analysis of data is covered in Chapter 4 at a level that is more complete than that found in other introductory analytical textbooks. Methods for calibrating equipment, standardizing methods, and linear regression are gathered together in Chapter 5. Chapter 6 provides an introduction to equilibrium chemistry, stressing both the rigorous solution to equilibrium problems, and the use of semi-quantitative approaches, such as ladder diagrams. The importance of collecting the right sample, and methods for separating analytes and interferents are covered in Chapter 7.

Chapters 8–13 cover the major areas of analysis, including gravimetry (Chapter 8), titrimetry (Chapter 9), spectroscopy (Chapter 10), electrochemistry (Chapter 11), chromatography and electrophoresis (Chapter 12), and kinetic methods (Chapter 13). Related techniques, such as acid–base titrimetry and redox titrimetry, or potentiometry and voltammetry, are gathered together in single chapters. Combining related techniques together encourages students to see the similarities between methods, rather than focusing on their differences. The first technique presented in each chapter is generally that which is most commonly covered in the introductory course.

Finally, the textbook concludes with two chapters discussing the design and maintenance of analytical methods, two topics of importance to analytical chemists. Chapter 14 considers the development of an analytical method, including its optimization, verification, and validation. Quality control and quality assessment are discussed in Chapter 15.

Acknowledgments

Before beginning an academic career I was, of course, a student. My interest in chemistry and teaching was nurtured by many fine teachers at Westtown Friends School, Knox College, and the University of North Carolina at Chapel Hill; their collective influence continues to bear fruit. In particular, I wish to recognize David MacInnes, Alan Hiebert, Robert Kooser, and Richard Linton.

I have been fortunate to work with many fine colleagues during my nearly 17 years of teaching undergraduate chemistry at Stockton State College and DePauw University. I am particularly grateful for the friendship and guidance provided by Jon Griffiths and Ed Paul during my four years at Stockton State College. At DePauw University, Jim George and Bryan Hanson have willingly shared their ideas about teaching, while patiently listening to mine.

Approximately 300 students have joined me in thinking and learning about analytical chemistry; their questions and comments helped guide the development of this textbook. I realize that working without a formal textbook has been frustrating and awkward; all the more reason why I appreciate their effort and hard work.

The following individuals reviewed portions of this textbook at various stages during its development.

David Ballantine
Northern Illinois University

John E. Bauer
Illinois State University

Ali Bazzi
University of Michigan–Dearborn

Steven D. Brown
University of Delaware

Wendy Clevenger
University of Tennessee–Chattanooga

Cathy Cobb
Augusta State University

Paul Flowers
University of North Carolina–Pembroke

Nancy Gordon
University of Southern Maine

Virginia M. Indivero
Swarthmore College

Michael Janusa
Nicholls State University

J. David Jenkins
Georgia Southern University

Richard S. Mitchell
Arkansas State University

George A. Pearse, Jr.
Le Moyne College

Gary Rayson
New Mexico State University

David Redfield
NW Nazarene University

Vincent Remcho
West Virginia University

Jeanette K. Rice
Georgia Southern University

Martin W. Rowe
Texas A&M University

Alexander Scheeline
University of Illinois

James D. Stuart
University of Connecticut

Thomas J. Wenzel
Bates College

David Zax
Cornell University

I am particularly grateful for their detailed written comments and suggestions for improving the manuscript. Much of what is good in the final manuscript is the result of their interest and ideas. George Foy (*York College of Pennsylvania*), John McBride (*Hofstra University*), and David Karpovich (*Saginaw Valley State University*) checked the accuracy of problems in the textbook. Gary Kinsel (*University of Texas at Arlington*) reviewed the page proofs and provided additional suggestions.

This project began in the summer of 1992 with the support of a course development grant from DePauw University's Faculty Development Fund. Additional financial support from DePauw University's Presidential Discretionary Fund also is acknowledged. Portions of the first draft were written during a sabbatical leave in the Fall semester of the 1993/94 academic year. A Fisher Fellowship provided release time during the Fall 1995 semester to complete the manuscript's second draft.

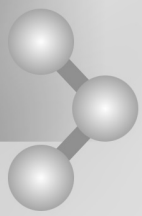
Alltech and Associates (Deerfield, IL) graciously provided permission to use the chromatograms in Chapter 12; the assistance of Jim Anderson, Vice-President, and Julia Poncher, Publications Director, is greatly appreciated. Fred Soster and Marilyn Culler, both of DePauw University, provided assistance with some of the photographs.

The editorial staff at McGraw-Hill has helped guide a novice through the process of developing this text. I am particularly thankful for the encouragement and confidence shown by Jim Smith, Publisher for Chemistry, and Kent Peterson, Sponsoring Editor for Chemistry. Shirley Oberbroeckling, Developmental Editor for Chemistry, and Jayne Klein, Senior Project Manager, patiently answered my questions and successfully guided me through the publishing process.

Finally, I would be remiss if I did not recognize the importance of my family's support and encouragement, particularly that of my parents. A very special thanks to my daughter, Devon, for gifts too numerous to detail.

How to Contact the Author

Writing this textbook has been an interesting (and exhausting) challenge. Despite my efforts, I am sure there are a few glitches, better examples, more interesting end-of-chapter problems, and better ways to think about some of the topics. I welcome your comments, suggestions, and data for interesting problems, which may be addressed to me at DePauw University, 602 S. College St., Greencastle, IN 46135, or electronically at harvey@depauw.edu.



Chapter 1

Introduction

Chemistry is the study of matter, including its composition, structure, physical properties, and reactivity. There are many approaches to studying chemistry, but, for convenience, we traditionally divide it into five fields: organic, inorganic, physical, biochemical, and analytical. Although this division is historical and arbitrary, as witnessed by the current interest in interdisciplinary areas such as bioanalytical and organometallic chemistry, these five fields remain the simplest division spanning the discipline of chemistry.

Training in each of these fields provides a unique perspective to the study of chemistry. Undergraduate chemistry courses and textbooks are more than a collection of facts; they are a kind of apprenticeship. In keeping with this spirit, this text introduces the field of analytical chemistry and the unique perspectives that analytical chemists bring to the study of chemistry.

1A What Is Analytical Chemistry?

*"Analytical chemistry is what analytical chemists do."**

We begin this section with a deceptively simple question. What is analytical chemistry? Like all fields of chemistry, analytical chemistry is too broad and active a discipline for us to easily or completely define in an introductory textbook. Instead, we will try to say a little about what analytical chemistry is, as well as a little about what analytical chemistry is not.

Analytical chemistry is often described as the area of chemistry responsible for characterizing the composition of matter, both qualitatively (what is present) and quantitatively (how much is present). This description is misleading. After all, almost all chemists routinely make qualitative or quantitative measurements. The argument has been made that analytical chemistry is not a separate branch of chemistry, but simply the application of chemical knowledge.[1] In fact, you probably have performed quantitative and qualitative analyses in other chemistry courses. For example, many introductory courses in chemistry include qualitative schemes for identifying inorganic ions and quantitative analyses involving titrations.

Unfortunately, this description ignores the unique perspective that analytical chemists bring to the study of chemistry. The craft of analytical chemistry is not in performing a routine analysis on a routine sample (which is more appropriately called chemical analysis), but in improving established methods, extending existing methods to new types of samples, and developing new methods for measuring chemical phenomena.[2]

Here's one example of this distinction between analytical chemistry and chemical analysis. Mining engineers evaluate the economic feasibility of extracting an ore by comparing the cost of removing the ore with the value of its contents. To estimate its value they analyze a sample of the ore. The challenge of developing and validating the method providing this information is the analytical chemist's responsibility. Once developed, the routine, daily application of the method becomes the job of the chemical analyst.

Another distinction between analytical chemistry and chemical analysis is that analytical chemists work to improve established methods. For example, several factors complicate the quantitative analysis of Ni^{2+} in ores, including the presence of a complex heterogeneous mixture of silicates and oxides, the low concentration of Ni^{2+} in ores, and the presence of other metals that may interfere in the analysis. Figure 1.1 is a schematic outline of one standard method in use during the late nineteenth century.[3] After dissolving a sample of the ore in a mixture of H_2SO_4 and HNO_3, trace metals that interfere with the analysis, such as Pb^{2+}, Cu^{2+} and Fe^{3+}, are removed by precipitation. Any cobalt and nickel in the sample are reduced to Co and Ni, isolated by filtration and weighed (point A). After dissolving the mixed solid, Co is isolated and weighed (point B). The amount of nickel in the ore sample is determined from the difference in the masses at points A and B.

$$\%\text{Ni} = \frac{\text{mass point A} - \text{mass point B}}{\text{mass sample}} \times 100$$

*Attributed to C. N. Reilley (1925–1981) on receipt of the 1965 Fisher Award in Analytical Chemistry. Reilley, who was a professor of chemistry at the University of North Carolina at Chapel Hill, was one of the most influential analytical chemists of the last half of the twentieth century.

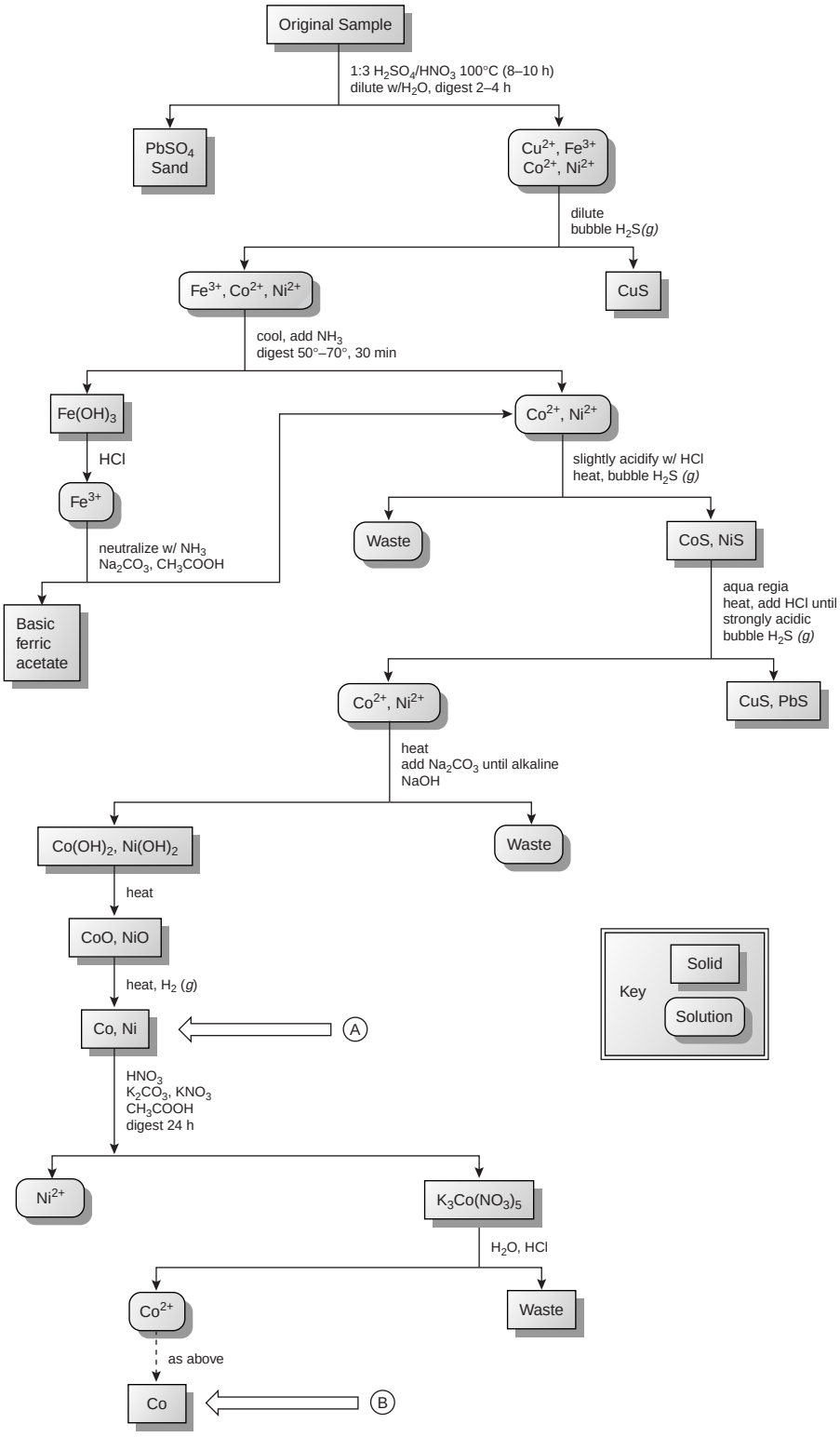


Figure 1.1
Analytical scheme outlined by Fresenius[3] for the gravimetric analysis of Ni in ores.

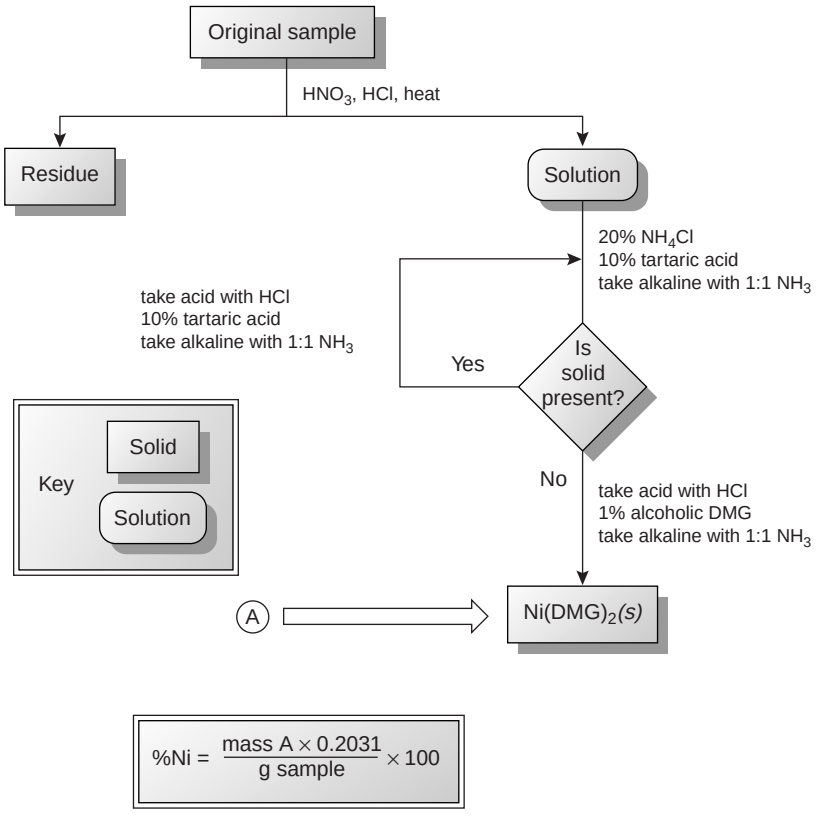


Figure 1.2

Analytical scheme outlined by Hillebrand and Lundell[4] for the gravimetric analysis of Ni in ores (DMG = dimethylgloxime). The factor of 0.2031 in the equation for %Ni accounts for the difference in the formula weights of $Ni(DMG)_2$ and Ni; see Chapter 8 for more details.

The combination of determining the mass of Ni^{2+} by difference, coupled with the need for many reactions and filtrations makes this procedure both time-consuming and difficult to perform accurately.

The development, in 1905, of dimethylgloxime (DMG), a reagent that selectively precipitates Ni^{2+} and Pd^{2+}, led to an improved analytical method for determining Ni^{2+} in ores.[4] As shown in Figure 1.2, the mass of Ni^{2+} is measured directly, requiring fewer manipulations and less time. By the 1970s, the standard method for the analysis of Ni^{2+} in ores progressed from precipitating $Ni(DMG)_2$ to flame atomic absorption spectrophotometry,[5] resulting in an even more rapid analysis. Current interest is directed toward using inductively coupled plasmas for determining trace metals in ores.

In summary, a more appropriate description of analytical chemistry is ". . . the science of inventing and applying the concepts, principles, and . . . strategies for measuring the characteristics of chemical systems and species."[6] Analytical chemists typically operate at the extreme edges of analysis, extending and improving the ability of all chemists to make meaningful measurements on smaller samples, on more complex samples, on shorter time scales, and on species present at lower concentrations. Throughout its history, analytical chemistry has provided many of the tools and methods necessary for research in the other four traditional areas of chemistry, as well as fostering multidisciplinary research in, to name a few, medicinal chemistry, clinical chemistry, toxicology, forensic chemistry, material science, geochemistry, and environmental chemistry.

You will come across numerous examples of qualitative and quantitative methods in this text, most of which are routine examples of chemical analysis. It is important to remember, however, that nonroutine problems prompted analytical chemists to develop these methods. Whenever possible, we will try to place these methods in their appropriate historical context. In addition, examples of current research problems in analytical chemistry are scattered throughout the text.

The next time you are in the library, look through a recent issue of an analytically oriented journal, such as *Analytical Chemistry.* Focus on the titles and abstracts of the research articles. Although you will not recognize all the terms and methods, you will begin to answer for yourself the question "What is analytical chemistry"?

1B The Analytical Perspective

Having noted that each field of chemistry brings a unique perspective to the study of chemistry, we now ask a second deceptively simple question. What is the "analytical perspective"? Many analytical chemists describe this perspective as an analytical approach to solving problems.[7] Although there are probably as many descriptions of the analytical approach as there are analytical chemists, it is convenient for our purposes to treat it as a five-step process:

1. Identify and define the problem.
2. Design the experimental procedure.
3. Conduct an experiment, and gather data.
4. Analyze the experimental data.
5. Propose a solution to the problem.

Figure 1.3 shows an outline of the analytical approach along with some important considerations at each step. Three general features of this approach deserve attention. First, steps 1 and 5 provide opportunities for analytical chemists to collaborate with individuals outside the realm of analytical chemistry. In fact, many problems on which analytical chemists work originate in other fields. Second, the analytical approach is not linear, but incorporates a "feedback loop" consisting of steps 2, 3, and 4, in which the outcome of one step may cause a reevaluation of the other two steps. Finally, the solution to one problem often suggests a new problem.

Analytical chemistry begins with a problem, examples of which include evaluating the amount of dust and soil ingested by children as an indicator of environmental exposure to particulate based pollutants, resolving contradictory evidence regarding the toxicity of perfluoro polymers during combustion, or developing rapid and sensitive detectors for chemical warfare agents.* At this point the analytical approach involves a collaboration between the analytical chemist and the individuals responsible for the problem. Together they decide what information is needed. It is also necessary for the analytical chemist to understand how the problem relates to broader research goals. The type of information needed and the problem's context are essential to designing an appropriate experimental procedure.

Designing an experimental procedure involves selecting an appropriate method of analysis based on established criteria, such as accuracy, precision, sensitivity, and detection limit; the urgency with which results are needed; the cost of a single analysis; the number of samples to be analyzed; and the amount of sample available for

*These examples are taken from a series of articles, entitled the "Analytical Approach," which has appeared as a regular feature in the journal *Analytical Chemistry* since 1974.

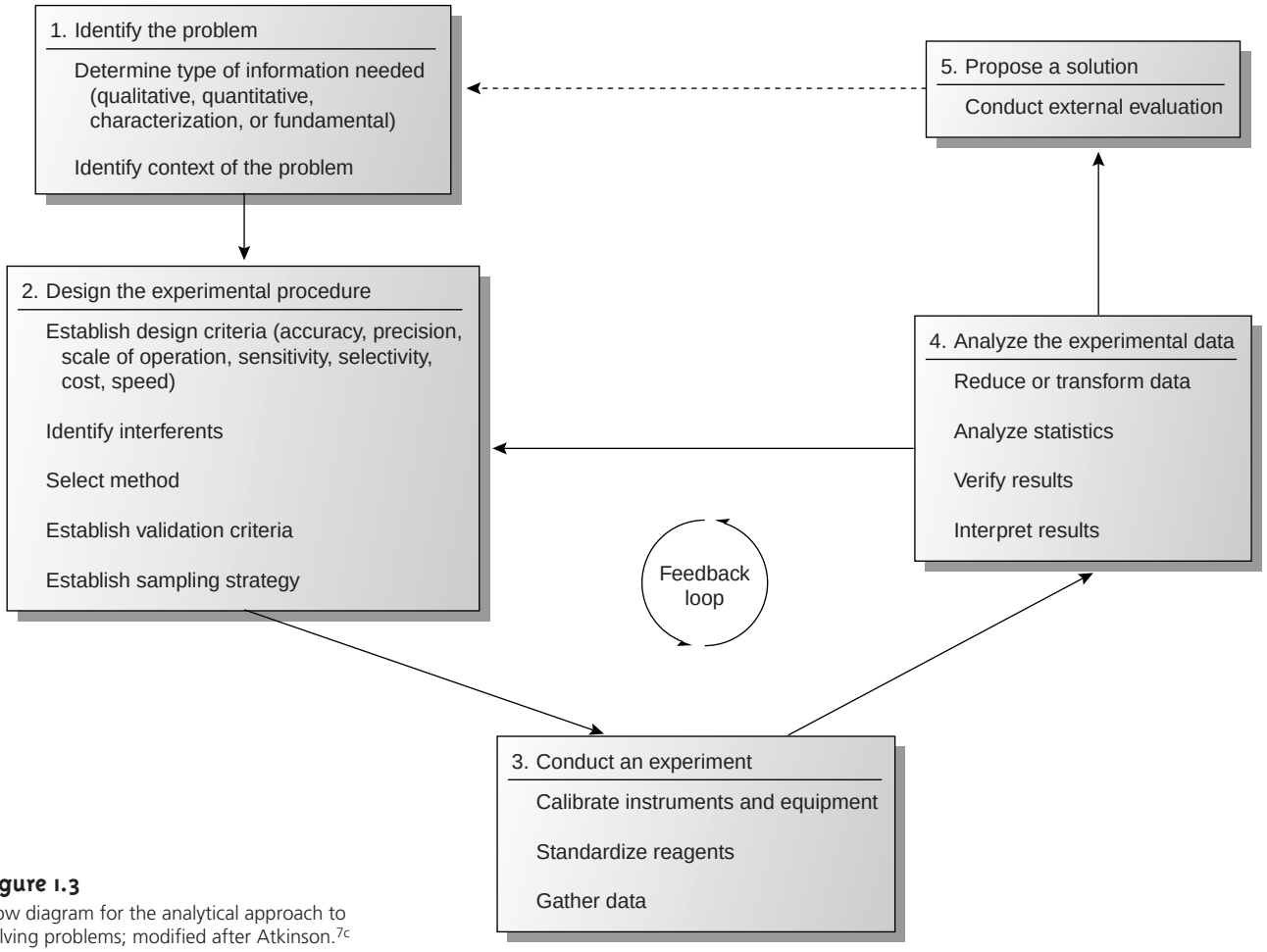


Figure 1.3
Flow diagram for the analytical approach to solving problems; modified after Atkinson.[7c]

analysis. Finding an appropriate balance between these parameters is frequently complicated by their interdependence. For example, improving the precision of an analysis may require a larger sample. Consideration is also given to collecting, storing, and preparing samples, and to whether chemical or physical interferences will affect the analysis. Finally, a good experimental procedure may still yield useless information if there is no method for validating the results.

The most visible part of the analytical approach occurs in the laboratory. As part of the validation process, appropriate chemical or physical standards are used to calibrate any equipment being used and any solutions whose concentrations must be known. The selected samples are then analyzed and the raw data recorded.

The raw data collected during the experiment are then analyzed. Frequently the data must be reduced or transformed to a more readily analyzable form. A statistical treatment of the data is used to evaluate the accuracy and precision of the analysis and to validate the procedure. These results are compared with the criteria established during the design of the experiment, and then the design is reconsidered, additional experimental trials are run, or a solution to the problem is proposed. When a solution is proposed, the results are subject to an external evaluation that may result in a new problem and the beginning of a new analytical cycle.

As an exercise, let's adapt this model of the analytical approach to a real problem. For our example, we will use the determination of the sources of airborne pollutant particles. A description of the problem can be found in the following article:

"Tracing Aerosol Pollutants with Rare Earth Isotopes" by
Ondov, J. M.; Kelly, W. R. *Anal. Chem.* **1991,** *63,* 691A–697A.

Before continuing, take some time to read the article, locating the discussions pertaining to each of the five steps outlined in Figure 1.3. In addition, consider the following questions:

1. What is the analytical problem?
2. What type of information is needed to solve the problem?
3. How will the solution to this problem be used?
4. What criteria were considered in designing the experimental procedure?
5. Were there any potential interferences that had to be eliminated? If so, how were they treated?
6. Is there a plan for validating the experimental method?
7. How were the samples collected?
8. Is there evidence that steps 2, 3, and 4 of the analytical approach are repeated more than once?
9. Was there a successful conclusion to the problem?

According to our model, the analytical approach begins with a problem. The motivation for this research was to develop a method for monitoring the transport of solid aerosol particulates following their release from a high-temperature combustion source. Because these particulates contain significant concentrations of toxic heavy metals and carcinogenic organic compounds, they represent a significant environmental hazard.

An aerosol is a suspension of either a solid or a liquid in a gas. Fog, for example, is a suspension of small liquid water droplets in air, and smoke is a suspension of small solid particulates in combustion gases. In both cases the liquid or solid particulates must be small enough to remain suspended in the gas for an extended time. Solid aerosol particulates, which are the focus of this problem, usually have micrometer or submicrometer diameters. Over time, solid particulates settle out from the gas, falling to the Earth's surface as dry deposition.

Existing methods for monitoring the transport of gases were inadequate for studying aerosols. To solve the problem, qualitative and quantitative information were needed to determine the sources of pollutants and their net contribution to the total dry deposition at a given location. Eventually the methods developed in this study could be used to evaluate models that estimate the contributions of point sources of pollution to the level of pollution at designated locations.

Following the movement of airborne pollutants requires a natural or artificial tracer (a species specific to the source of the airborne pollutants) that can be experimentally measured at sites distant from the source. Limitations placed on the tracer, therefore, governed the design of the experimental procedure. These limitations included cost, the need to detect small quantities of the tracer, and the absence of the tracer from other natural sources. In addition, aerosols are emitted from high-temperature combustion sources that produce an abundance of very reactive species. The tracer, therefore, had to be both thermally and chemically stable. On the basis of these criteria, rare earth isotopes, such as those of Nd, were selected as tracers. The choice of tracer, in turn, dictated the analytical method (thermal ionization mass spectrometry, or TIMS) for measuring the isotopic abundances of

Nd in samples. Unfortunately, mass spectrometry is not a selective technique. A mass spectrum provides information about the abundance of ions with a given mass. It cannot distinguish, however, between different ions with the same mass. Consequently, the choice of TIMS required developing a procedure for separating the tracer from the aerosol particulates.

Validating the final experimental protocol was accomplished by running a model study in which ^{148}Nd was released into the atmosphere from a 100-MW coal utility boiler. Samples were collected at 13 locations, all of which were 20 km from the source. Experimental results were compared with predictions determined by the rate at which the tracer was released and the known dispersion of the emissions.

Finally, the development of this procedure did not occur in a single, linear pass through the analytical approach. As research progressed, problems were encountered and modifications made, representing a cycle through steps 2, 3, and 4 of the analytical approach.

Others have pointed out, with justification, that the analytical approach outlined here is not unique to analytical chemistry, but is common to any aspect of science involving analysis.[8] Here, again, it helps to distinguish between a chemical analysis and analytical chemistry. For other analytically oriented scientists, such as physical chemists and physical organic chemists, the primary emphasis is on the problem, with the results of an analysis supporting larger research goals involving fundamental studies of chemical or physical processes. The essence of analytical chemistry, however, is in the second, third, and fourth steps of the analytical approach. Besides supporting broader research goals by developing and validating analytical methods, these methods also define the type and quality of information available to other research scientists. In some cases, the success of an analytical method may even suggest new research problems.

1C Common Analytical Problems

In Section 1A we indicated that analytical chemistry is more than a collection of qualitative and quantitative methods of analysis. Nevertheless, many problems on which analytical chemists work ultimately involve either a qualitative or quantitative measurement. Other problems may involve characterizing a sample's chemical or physical properties. Finally, many analytical chemists engage in fundamental studies of analytical methods. In this section we briefly discuss each of these four areas of analysis.

qualitative analysis
An analysis in which we determine the identity of the constituent species in a sample.

Many problems in analytical chemistry begin with the need to identify what is present in a sample. This is the scope of a **qualitative analysis,** examples of which include identifying the products of a chemical reaction, screening an athlete's urine for the presence of a performance-enhancing drug, or determining the spatial distribution of Pb on the surface of an airborne particulate. Much of the early work in analytical chemistry involved the development of simple chemical tests to identify the presence of inorganic ions and organic functional groups. The classical laboratory courses in inorganic and organic qualitative analysis,[9] still taught at some schools, are based on this work. Currently, most qualitative analyses use methods such as infrared spectroscopy, nuclear magnetic resonance, and mass spectrometry. These qualitative applications of identifying organic and inorganic compounds are covered adequately elsewhere in the undergraduate curriculum and, so, will receive no further consideration in this text.

Perhaps the most common type of problem encountered in the analytical lab is a **quantitative analysis.** Examples of typical quantitative analyses include the elemental analysis of a newly synthesized compound, measuring the concentration of glucose in blood, or determining the difference between the bulk and surface concentrations of Cr in steel. Much of the analytical work in clinical, pharmaceutical, environmental, and industrial labs involves developing new methods for determining the concentration of targeted species in complex samples. Most of the examples in this text come from the area of quantitative analysis.

quantitative analysis
An analysis in which we determine how much of a constituent species is present in a sample.

Another important area of analytical chemistry, which receives some attention in this text, is the development of new methods for characterizing physical and chemical properties. Determinations of chemical structure, equilibrium constants, particle size, and surface structure are examples of a **characterization analysis.**

characterization analysis
An analysis in which we evaluate a sample's chemical or physical properties.

The purpose of a qualitative, quantitative, and characterization analysis is to solve a problem associated with a sample. A **fundamental analysis,** on the other hand, is directed toward improving the experimental methods used in the other areas of analytical chemistry. Extending and improving the theory on which a method is based, studying a method's limitations, and designing new and modifying old methods are examples of fundamental studies in analytical chemistry.

fundamental analysis
An analysis whose purpose is to improve an analytical method's capabilities.

1D KEY TERMS

characterization analysis (*p. 9*) qualitative analysis (*p. 8*) quantitative analysis (*p. 9*)
fundamental analysis (*p. 9*)

1E SUMMARY

Analytical chemists work to improve the ability of all chemists to make meaningful measurements. Chemists working in medicinal chemistry, clinical chemistry, forensic chemistry, and environmental chemistry, as well as the more traditional areas of chemistry, need better tools for analyzing materials. The need to work with smaller quantities of material, with more complex materials, with processes occurring on shorter time scales, and with species present at lower concentrations challenges analytical chemists to improve existing analytical methods and to develop new analytical techniques.

Typical problems on which analytical chemists work include qualitative analyses (what is present?), quantitative analyses (how much is present?), characterization analyses (what are the material's chemical and physical properties?), and fundamental analyses (how does this method work and how can it be improved?).

1F PROBLEMS

1. For each of the following problems indicate whether its solution requires a qualitative, quantitative, characterization, or fundamental study. More than one type of analysis may be appropriate for some problems.
   a. A hazardous-waste disposal site is believed to be leaking contaminants into the local groundwater.
   b. An art museum is concerned that a recent acquisition is a forgery.
   c. A more reliable method is needed by airport security for detecting the presence of explosive materials in luggage.
   d. The structure of a newly discovered virus needs to be determined.
   e. A new visual indicator is needed for an acid–base titration.
   f. A new law requires a method for evaluating whether automobiles are emitting too much carbon monoxide.
2. Read a recent article from the column "Analytical Approach," published in *Analytical Chemistry,* or an article assigned by your instructor, and write an essay summarizing the nature of the problem and how it was solved. As a guide, refer back to Figure 1.3 for one model of the analytical approach.

1G SUGGESTED READINGS

The role of analytical chemistry within the broader discipline of chemistry has been discussed by many prominent analytical chemists. Several notable examples follow.

Baiulescu, G. E.; Patroescu, C.; Chalmers, R. A. *Education and Teaching in Analytical Chemistry.* Ellis Horwood: Chichester, 1982.

Hieftje, G. M. "The Two Sides of Analytical Chemistry," *Anal. Chem.* **1985,** *57,* 256A–267A.

Kissinger, P. T. "Analytical Chemistry—What is It? Who Needs It? Why Teach It?" *Trends Anal. Chem.* **1992,** *11,* 54–57.

Laitinen, H. A. "Analytical Chemistry in a Changing World," *Anal. Chem.* **1980,** *52,* 605A–609A.

Laitinen, H. A. "History of Analytical Chemistry in the U.S.A.," *Talanta* **1989,** *36,* 1–9.

Laitinen, H. A.; Ewing, G. (eds). *A History of Analytical Chemistry.* The Division of Analytical Chemistry of the American Chemical Society: Washington, D.C., 1972.

McLafferty, F. W. "Analytical Chemistry: Historic and Modern," *Acc. Chem. Res.* **1990,** *23,* 63–64.

Mottola, H. A. "The Interdisciplinary and Multidisciplinary Nature of Contemporary Analytical Chemistry and Its Core Components," *Anal. Chim. Acta* **1991,** *242,* 1–3.

Tyson, J. *Analysis: What Analytical Chemists Do.* Royal Society of Chemistry: Cambridge, England, 1988.

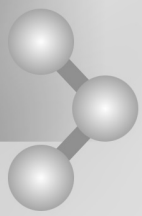
Several journals are dedicated to publishing broadly in the field of analytical chemistry, including *Analytical Chemistry, Analytica Chimica Acta, Analyst,* and *Talanta.* Other journals, too numerous to list, are dedicated to single areas of analytical chemistry.

Current research in the areas of quantitative analysis, qualitative analysis, and characterization analysis are reviewed biannually (odd-numbered years) in *Analytical Chemistry's* "Application Reviews."

Current research on fundamental developments in analytical chemistry are reviewed biannually (even-numbered years) in *Analytical Chemistry's* "Fundamental Reviews."

1H REFERENCES

1. Ravey, M. *Spectroscopy* **1990,** *5(7),* 11.
2. de Haseth, J. *Spectroscopy* **1990,** *5(7),* 11.
3. Fresenius, C. R. *A System of Instruction in Quantitative Chemical Analysis.* John Wiley and Sons: New York, 1881.
4. Hillebrand, W. F.; Lundell, G. E. F. *Applied Inorganic Analysis,* John Wiley and Sons: New York, 1953.
5. Van Loon, J. C. *Analytical Atomic Absorption Spectroscopy.* Academic Press: New York, 1980.
6. Murray, R. W. *Anal. Chem.* **1991,** *63,* 271A.
7. For several different viewpoints see (a) Beilby, A. L. *J. Chem. Educ.* **1970,** *47,* 237–238; (b) Lucchesi, C. A. *Am. Lab.* **1980,** *October,* 113–119; (c) Atkinson, G. F. *J. Chem. Educ.* **1982,** *59,* 201–202; (d) Pardue, H. L.; Woo, J. *J. Chem. Educ.* **1984,** *61,* 409–412; (e) Guarnieri, M. *J. Chem. Educ.* **1988,** *65,* 201–203; (f) de Haseth, J. *Spectroscopy* **1990,** *5,* 20–21; (g) Strobel, H. A. *Am. Lab.* **1990,** *October,* 17–24.
8. Hieftje, G. M. *Am. Lab.* **1993,** *October,* 53–61.
9. See, for example, the following laboratory texts: (a) Sorum, C. H.; Lagowski, J. J. *Introduction to Semimicro Qualitative Analysis,* 5th ed. Prentice-Hall: Englewood Cliffs, NJ, 1977.; (b) Shriner, R. L.; Fuson, R. C.; Curtin, D. Y. *The Systematic Identification of Organic Compounds,* 5th ed. John Wiley and Sons: New York, 1964.



Chapter 2

Basic Tools of Analytical Chemistry

In the chapters that follow we will learn about the specifics of analytical chemistry. In the process we will ask and answer questions such as "How do we treat experimental data?" "How do we ensure that our results are accurate?" "How do we obtain a representative sample?" and "How do we select an appropriate analytical technique?" Before we look more closely at these and other questions, we will first review some basic numerical and experimental tools of importance to analytical chemists.

2A Numbers in Analytical Chemistry

Analytical chemistry is inherently a quantitative science. Whether determining the concentration of a species in a solution, evaluating an equilibrium constant, measuring a reaction rate, or drawing a correlation between a compound's structure and its reactivity, analytical chemists make measurements and perform calculations. In this section we briefly review several important topics involving the use of numbers in analytical chemistry.

2A.1 Fundamental Units of Measure

Imagine that you find the following instructions in a laboratory procedure: "Transfer 1.5 of your sample to a 100 volumetric flask, and dilute to volume." How do you do this? Clearly these instructions are incomplete since the units of measurement are not stated. Compare this with a complete instruction: "Transfer 1.5 g of your sample to a 100-mL volumetric flask, and dilute to volume." This is an instruction that you can easily follow.

Measurements usually consist of a unit and a number expressing the quantity of that unit. Unfortunately, many different units may be used to express the same physical measurement. For example, the mass of a sample weighing 1.5 g also may be expressed as 0.0033 lb or 0.053 oz. For consistency, and to avoid confusion, scientists use a common set of fundamental units, several of which are listed in Table 2.1. These units are called **SI units** after the *Système International d'Unités.* Other measurements are defined using these fundamental SI units. For example, we measure the quantity of heat produced during a chemical reaction in joules, (J), where

SI units
Stands for *Système International d'Unités.* These are the internationally agreed on units for measurements.

$$1\ \text{J} = 1\frac{\text{m}^2\text{kg}}{\text{s}^2}$$

Table 2.2 provides a list of other important derived SI units, as well as a few commonly used non-SI units.

Chemists frequently work with measurements that are very large or very small. A mole, for example, contains 602,213,670,000,000,000,000,000 particles, and some analytical techniques can detect as little as 0.000000000000001 g of a compound. For simplicity, we express these measurements using **scientific notation;** thus, a mole contains 6.0221367×10^{23} particles, and the stated mass is 1×10^{-15} g. Sometimes it is preferable to express measurements without the exponential term, replacing it with a prefix. A mass of 1×10^{-15} g is the same as 1 femtogram. Table 2.3 lists other common prefixes.

scientific notation
A shorthand method for expressing very large or very small numbers by indicating powers of ten; for example, 1000 is 1×10^3.

Table 2.1 Fundamental SI Units

Measurement	Unit	Symbol
mass	kilogram	kg
volume	liter	L
distance	meter	m
temperature	kelvin	K
time	second	s
current	ampere	A
amount of substance	mole	mol

Table 2.2 Other SI and Non-SI Units

Measurement	Unit	Symbol	Equivalent SI units
length	angstrom	Å	$1\ Å = 1 \times 10^{-10}\ m$
force	newton	N	$1\ N = 1\ m \cdot kg/s^2$
pressure	pascal	Pa	$1\ Pa = 1\ N/m^2 = 1\ kg/(m \cdot s^2)$
	atmosphere	atm	1 atm = 101,325 Pa
energy, work, heat	joule	J	$1\ J = 1\ N \cdot m = 1\ m^2 \cdot kg/s^2$
power	watt	W	$1\ W = 1\ J/s = 1\ m^2 \cdot kg/s^3$
charge	coulomb	C	$1\ C = 1\ A \cdot s$
potential	volt	V	$1\ V = 1\ W/A = 1\ m^2 \cdot kg/(s^3 \cdot A)$
temperature	degree Celsius	°C	°C = K – 273.15
	degree Fahrenheit	°F	°F = 1.8(K – 273.15) + 32

Table 2.3 Common Prefixes for Exponential Notation

Exponential	Prefix	Symbol
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a

2A.2 Significant Figures

Recording a measurement provides information about both its magnitude and uncertainty. For example, if we weigh a sample on a balance and record its mass as 1.2637 g, we assume that all digits, except the last, are known exactly. We assume that the last digit has an uncertainty of at least ±1, giving an absolute uncertainty of at least ±0.0001 g, or a relative uncertainty of at least

$$\frac{\pm 0.0001\ g}{1.2637\ g} \times 100 = \pm 0.0079\%$$

Significant figures are a reflection of a measurement's uncertainty. The number of significant figures is equal to the number of digits in the measurement, with the exception that a zero (0) used to fix the location of a decimal point is not considered significant. This definition can be ambiguous. For example, how many significant figures are in the number 100? If measured to the nearest hundred, then there is one significant figure. If measured to the nearest ten, however, then two

significant figures
The digits in a measured quantity, including all digits known exactly and one digit (the last) whose quantity is uncertain.

significant figures are included. To avoid ambiguity we use scientific notation. Thus, 1×10^2 has one significant figure, whereas 1.0×10^2 has two significant figures.

For measurements using logarithms, such as pH, the number of significant figures is equal to the number of digits to the right of the decimal, including all zeros. Digits to the left of the decimal are not included as significant figures since they only indicate the power of 10. A pH of 2.45, therefore, contains two significant figures.

Exact numbers, such as the stoichiometric coefficients in a chemical formula or reaction, and unit conversion factors, have an infinite number of significant figures. A mole of $CaCl_2$, for example, contains exactly two moles of chloride and one mole of calcium. In the equality

$$1000 \text{ mL} = 1 \text{ L}$$

both numbers have an infinite number of significant figures.

Recording a measurement to the correct number of significant figures is important because it tells others about how precisely you made your measurement. For example, suppose you weigh an object on a balance capable of measuring mass to the nearest ±0.1 mg, but record its mass as 1.762 g instead of 1.7620 g. By failing to record the trailing zero, which is a significant figure, you suggest to others that the mass was determined using a balance capable of weighing to only the nearest ±1 mg. Similarly, a buret with scale markings every 0.1 mL can be read to the nearest ±0.01 mL. The digit in the hundredth's place is the least significant figure since we must estimate its value. Reporting a volume of 12.241 mL implies that your buret's scale is more precise than it actually is, with divisions every 0.01 mL.

Significant figures are also important because they guide us in reporting the result of an analysis. When using a measurement in a calculation, the result of that calculation can never be more certain than that measurement's uncertainty. Simply put, the result of an analysis can never be more certain than the least certain measurement included in the analysis.

As a general rule, mathematical operations involving addition and subtraction are carried out to the last digit that is significant for all numbers included in the calculation. Thus, the sum of 135.621, 0.33, and 21.2163 is 157.17 since the last digit that is significant for all three numbers is in the hundredth's place.

$$135.6\underline{2}1 + 0.3\underline{3} + 21.2\underline{1}63 = 157.1673 = 157.1\underline{7}$$

When multiplying and dividing, the general rule is that the answer contains the same number of significant figures as that number in the calculation having the fewest significant figures. Thus,

$$\frac{22.91 \times 0.\underline{152}}{16.302} = 0.21361 = 0.\underline{214}$$

It is important to remember, however, that these rules are generalizations. What is conserved is not the number of significant figures, but absolute uncertainty when adding or subtracting, and relative uncertainty when multiplying or dividing. For example, the following calculation reports the answer to the correct number of significant figures, even though it violates the general rules outlined earlier.

$$\frac{101}{99} = 1.02$$

Since the relative uncertainty in both measurements is roughly 1% (101 ±1, 99 ±1), the relative uncertainty in the final answer also must be roughly 1%. Reporting the answer to only two significant figures (1.0), as required by the general rules, implies a relative uncertainty of 10%. The correct answer, with three significant figures, yields the expected relative uncertainty. Chapter 4 presents a more thorough treatment of uncertainty and its importance in reporting the results of an analysis.

Finally, to avoid "round-off" errors in calculations, it is a good idea to retain at least one extra significant figure throughout the calculation. This is the practice adopted in this textbook. Better yet, invest in a good scientific calculator that allows you to perform lengthy calculations without recording intermediate values. When the calculation is complete, the final answer can be rounded to the correct number of significant figures using the following simple rules.

1. Retain the least significant figure if it and the digits that follow are less than halfway to the next higher digit; thus, rounding 12.442 to the nearest tenth gives 12.4 since 0.442 is less than halfway between 0.400 and 0.500.
2. Increase the least significant figure by 1 if it and the digits that follow are more than halfway to the next higher digit; thus, rounding 12.476 to the nearest tenth gives 12.5 since 0.476 is more than halfway between 0.400 and 0.500.
3. If the least significant figure and the digits that follow are exactly halfway to the next higher digit, then round the least significant figure to the nearest even number; thus, rounding 12.450 to the nearest tenth gives 12.4, but rounding 12.550 to the nearest tenth gives 12.6. Rounding in this manner prevents us from introducing a bias by always rounding up or down.

2B Units for Expressing Concentration

Concentration is a general measurement unit stating the amount of solute present in a known amount of solution

concentration
An expression stating the relative amount of solute per unit volume or unit mass of solution.

$$\text{Concentration} = \frac{\text{amount of solute}}{\text{amount of solution}} \qquad \textbf{2.1}$$

Although the terms "solute" and "solution" are often associated with liquid samples, they can be extended to gas-phase and solid-phase samples as well. The actual units for reporting concentration depend on how the amounts of solute and solution are measured. Table 2.4 lists the most common units of concentration.

2B.1 Molarity and Formality

Both molarity and formality express concentration as moles of solute per liter of solution. There is, however, a subtle difference between molarity and formality. **Molarity** is the concentration of a particular chemical species in solution. **Formality,** on the other hand, is a substance's total concentration in solution without regard to its specific chemical form. There is no difference between a substance's molarity and formality if it dissolves without dissociating into ions. The molar concentration of a solution of glucose, for example, is the same as its formality.

molarity
The number of moles of solute per liter of solution (M).

formality
The number of moles of solute, regardless of chemical form, per liter of solution (F).

For substances that ionize in solution, such as NaCl, molarity and formality are different. For example, dissolving 0.1 mol of NaCl in 1 L of water gives a solution containing 0.1 mol of Na^+ and 0.1 mol of Cl^-. The molarity of NaCl, therefore, is zero since there is essentially no undissociated NaCl in solution. The solution,

Table 2.4 Common Units for Reporting Concentration

Name	Units[a]	Symbol
molarity	$\frac{\text{moles solute}}{\text{liters solution}}$	M
formality	$\frac{\text{number FWs solute}}{\text{liters solution}}$	F
normality	$\frac{\text{number EWs solute}}{\text{liters solution}}$	N
molality	$\frac{\text{moles solute}}{\text{kg solvent}}$	*m*
weight %	$\frac{\text{g solute}}{\text{100 g solution}}$	% w/w
volume %	$\frac{\text{mL solute}}{\text{100 mL solution}}$	% v/v
weight-to-volume %	$\frac{\text{g solute}}{\text{100 mL solution}}$	% w/v
parts per million	$\frac{\text{g solute}}{10^6\ \text{g solution}}$	ppm
parts per billion	$\frac{\text{g solute}}{10^9\ \text{g solution}}$	ppb

[a]FW = formula weight; EW = equivalent weight.

instead, is 0.1 M in Na^+ and 0.1 M in Cl^-. The formality of NaCl, however, is 0.1 F because it represents the total amount of NaCl in solution. The rigorous definition of molarity, for better or worse, is largely ignored in the current literature, as it is in this text. When we state that a solution is 0.1 M NaCl we understand it to consist of Na^+ and Cl^- ions. The unit of formality is used only when it provides a clearer description of solution chemistry.

Molar concentrations are used so frequently that a symbolic notation is often used to simplify its expression in equations and writing. The use of square brackets around a species indicates that we are referring to that species' molar concentration. Thus, $[Na^+]$ is read as the "molar concentration of sodium ions."

2B.2 Normality

Normality is an older unit of concentration that, although once commonly used, is frequently ignored in today's laboratories. Normality is still used in some handbooks of analytical methods, and, for this reason, it is helpful to understand its meaning. For example, normality is the concentration unit used in *Standard Methods for the Examination of Water and Wastewater*,[1] a commonly used source of analytical methods for environmental laboratories.

normality
The number of equivalents of solute per liter of solution (N).

Normality makes use of the chemical equivalent, which is the amount of one chemical species reacting stoichiometrically with another chemical species. Note that this definition makes an equivalent, and thus normality, a function of the chemical reaction in which the species participates. Although a solution of H_2SO_4 has a fixed molarity, its normality depends on how it reacts.

The number of **equivalents,** *n,* is based on a reaction unit, which is that part of a chemical species involved in a reaction. In a precipitation reaction, for example, the reaction unit is the charge of the cation or anion involved in the reaction; thus for the reaction

equivalent
The moles of a species that can donate one reaction unit.

$$Pb^{2+}(aq) + 2I^{-}(aq) \rightleftharpoons PbI_2(s)$$

$n = 2$ for Pb^{2+} and $n = 1$ for I^-. In an acid–base reaction, the reaction unit is the number of H^+ ions donated by an acid or accepted by a base. For the reaction between sulfuric acid and ammonia

$$H_2SO_4(aq) + 2NH_3(aq) \rightleftharpoons 2NH_4^{+}(aq) + SO_4^{2-}(aq)$$

we find that $n = 2$ for H_2SO_4 and $n = 1$ for NH_3. For a complexation reaction, the reaction unit is the number of electron pairs that can be accepted by the metal or donated by the ligand. In the reaction between Ag^+ and NH_3

$$Ag^{+}(aq) + 2NH_3(aq) \rightleftharpoons Ag(NH_3)_2^{+}(aq)$$

the value of n for Ag^+ is 2 and that for NH_3 is 1. Finally, in an oxidation–reduction reaction the reaction unit is the number of electrons released by the reducing agent or accepted by the oxidizing agent; thus, for the reaction

$$2Fe^{3+}(aq) + Sn^{2+}(aq) \rightleftharpoons Sn^{4+}(aq) + 2Fe^{2+}(aq)$$

$n = 1$ for Fe^{3+} and $n = 2$ for Sn^{2+}. Clearly, determining the number of equivalents for a chemical species requires an understanding of how it reacts.

Normality is the number of **equivalent weights** (EW) per unit volume and, like formality, is independent of speciation. An equivalent weight is defined as the ratio of a chemical species' **formula weight** (FW) to the number of its equivalents

equivalent weight
The mass of a compound containing one equivalent (EW).

formula weight
The mass of a compound containing one mole (FW).

$$EW = \frac{FW}{n}$$

Consequently, the following simple relationship exists between normality and molarity.

$$N = n \times M$$

Example 2.1 illustrates the relationship among chemical reactivity, equivalent weight, and normality.

Example

EXAMPLE 2.1

Calculate the equivalent weight and normality for a solution of 6.0 M H_3PO_4 given the following reactions:

(a) $H_3PO_4(aq) + 3OH^{-}(aq) \rightleftharpoons PO_4^{3-}(aq) + 3H_2O(\ell)$

(b) $H_3PO_4(aq) + 2NH_3(aq) \rightleftharpoons HPO_4^{2-}(aq) + 2NH_4^{+}(aq)$

(c) $H_3PO_4(aq) + F^{-}(aq) \rightleftharpoons H_2PO_4^{-}(aq) + HF(aq)$

SOLUTION

For phosphoric acid, the number of equivalents is the number of H^+ ions donated to the base. For the reactions in (a), (b), and (c) the number of equivalents are 3, 2, and 1, respectively. Thus, the calculated equivalent weights and normalities are

(a) $EW = \frac{FW}{n} = \frac{97.994}{3} = 32.665$ $\quad N = n \times M = 3 \times 6.0 = 18\ N$

(b) $EW = \frac{FW}{n} = \frac{97.994}{2} = 48.997$ $\quad N = n \times M = 2 \times 6.0 = 12\ N$

(c) $EW = \frac{FW}{n} = \frac{97.994}{1} = 97.994$ $\quad N = n \times M = 1 \times 6.0 = 6.0\ N$

2B.3 Molality

molality
The number of moles of solute per kilogram of solvent (m).

Molality is used in thermodynamic calculations where a temperature independent unit of concentration is needed. Molarity, formality and normality are based on the volume of solution in which the solute is dissolved. Since density is a temperature dependent property a solution's volume, and thus its molar, formal and normal concentrations, will change as a function of its temperature. By using the solvent's mass in place of its volume, the resulting concentration becomes independent of temperature.

2B.4 Weight, Volume, and Weight-to-Volume Ratios

weight percent
Grams of solute per 100 g of solution. (% w/w).

volume percent
Milliliters of solute per 100 mL of solution (% v/v).

weight-to-volume percent
Grams of solute per 100 mL of solution (% w/v).

parts per million
Micrograms of solute per gram of solution; for aqueous solutions the units are often expressed as milligrams of solute per liter of solution (ppm).

parts per billion
Nanograms of solute per gram of solution; for aqueous solutions the units are often expressed as micrograms of solute per liter of solution (ppb).

Weight percent (% w/w), **volume percent** (% v/v) and **weight-to-volume percent** (% w/v) express concentration as units of solute per 100 units of sample. A solution in which a solute has a concentration of 23% w/v contains 23 g of solute per 100 mL of solution.

Parts per million (ppm) and **parts per billion** (ppb) are mass ratios of grams of solute to one million or one billion grams of sample, respectively. For example, a steel that is 450 ppm in Mn contains 450 μg of Mn for every gram of steel. If we approximate the density of an aqueous solution as 1.00 g/mL, then solution concentrations can be expressed in parts per million or parts per billion using the following relationships.

$$ppm = \frac{mg}{liter} = \frac{\mu g}{mL}$$

$$ppb = \frac{\mu g}{liter} = \frac{ng}{mL}$$

For gases a part per million usually is a volume ratio. Thus, a helium concentration of 6.3 ppm means that one liter of air contains 6.3 μL of He.

2B.5 Converting Between Concentration Units

The units of concentration most frequently encountered in analytical chemistry are molarity, weight percent, volume percent, weight-to-volume percent, parts per million, and parts per billion. By recognizing the general definition of concentration given in equation 2.1, it is easy to convert between concentration units.

Example

EXAMPLE 2.2

A concentrated solution of aqueous ammonia is 28.0% w/w NH_3 and has a density of 0.899 g/mL. What is the molar concentration of NH_3 in this solution?

SOLUTION

$$\frac{28.0\ g\ NH_3}{100\ g\ solution} \times \frac{0.899\ g\ solution}{mL\ solution} \times \frac{1\ mole\ NH_3}{17.04\ g\ NH_3} \times \frac{1000\ mL}{liter} = 14.8\ M$$

Example

EXAMPLE 2.3

The maximum allowed concentration of chloride in a municipal drinking water supply is 2.50×10^2 ppm Cl^-. When the supply of water exceeds this limit, it often has a distinctive salty taste. What is this concentration in moles Cl^-/liter?

SOLUTION

$$\frac{2.50 \times 10^2 \text{ mg } Cl^-}{\text{L}} \times \frac{1 \text{ g}}{1000 \text{ mg}} \times \frac{1 \text{ mole } Cl^-}{35.453 \text{ g } Cl^-} = 7.05 \times 10^{-3} \text{ M}$$

2B.6 p-Functions

Sometimes it is inconvenient to use the concentration units in Table 2.4. For example, during a reaction a reactant's concentration may change by many orders of magnitude. If we are interested in viewing the progress of the reaction graphically, we might wish to plot the reactant's concentration as a function of time or as a function of the volume of a reagent being added to the reaction. Such is the case in Figure 2.1, where the molar concentration of H^+ is plotted (y-axis on left side of figure) as a function of the volume of NaOH added to a solution of HCl. The initial $[H^+]$ is 0.10 M, and its concentration after adding 75 mL of NaOH is 5.0×10^{-13} M. We can easily follow changes in the $[H^+]$ over the first 14 additions of NaOH. For the last ten additions of NaOH, however, changes in the $[H^+]$ are too small to be seen.

When working with concentrations that span many orders of magnitude, it is often more convenient to express the concentration as a **p-function.** The p-function of a number X is written as pX and is defined as

p-function
A function of the form pX, where pX = -log(X).

$$pX = -\log(X)$$

Thus, the pH of a solution that is 0.10 M H^+ is

$$pH = -\log[H^+] = -\log(0.10) = 1.00$$

and the pH of 5.0×10^{-13} M H^+ is

$$pH = -\log[H^+] = -\log(5.0 \times 10^{-13}) = 12.30$$

Figure 2.1 shows how plotting pH in place of $[H^+]$ provides more detail about how the concentration of H^+ changes following the addition of NaOH.

Example

EXAMPLE 2.4

What is pNa for a solution of 1.76×10^{-3} M Na_3PO_4?

SOLUTION

Since each mole of Na_3PO_4 contains three moles of Na^+, the concentration of Na^+ is

$$[Na^+] = \frac{3 \text{ mol } Na^+}{\text{mol } Na_3PO_4} \times 1.76 \times 10^{-3} \text{ M} = 5.28 \times 10^{-3} \text{ M}$$

and pNa is

$$pNa = -\log[Na^+] = -\log(5.28 \times 10^{-3}) = 2.277$$

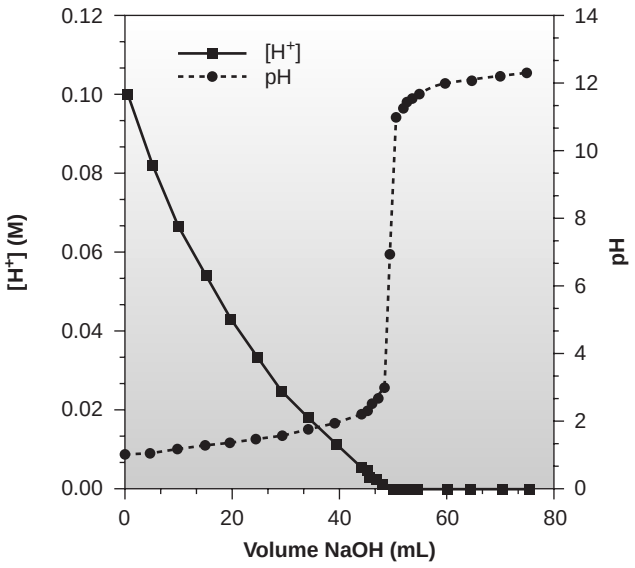


Figure 2.1
Graph of $[H^+]$ versus volume of NaOH and pH versus volume of NaOH for the reaction of 0.10 M HCl with 0.10 M NaOH.

Example

EXAMPLE 2.5

What is the $[H^+]$ in a solution that has a pH of 5.16?

SOLUTION

The concentration of H^+ is

$$pH = -\log[H^+] = 5.16$$

$$\log[H^+] = -5.16$$

$$[H^+] = \text{antilog}(-5.16) = 10^{-5.16} = 6.9 \times 10^{-6}\ \text{M}$$

2C Stoichiometric Calculations

A balanced chemical reaction indicates the quantitative relationships between the moles of reactants and products. These stoichiometric relationships provide the basis for many analytical calculations. Consider, for example, the problem of determining the amount of oxalic acid, $H_2C_2O_4$, in rhubarb. One method for this analysis uses the following reaction in which we oxidize oxalic acid to CO_2.

$$2Fe^{3+}(aq) + H_2C_2O_4(aq) + 2H_2O(\ell) \rightarrow 2Fe^{2+}(aq) + 2CO_2(g) + 2H_3O^+(aq) \quad \textbf{2.2}$$

The balanced chemical reaction provides the stoichiometric relationship between the moles of Fe^{3+} used and the moles of oxalic acid in the sample being analyzed—specifically, one mole of oxalic acid reacts with two moles of Fe^{3+}. As shown in Example 2.6, the balanced chemical reaction can be used to determine the amount of oxalic acid in a sample, provided that information about the number of moles of Fe^{3+} is known.

Example

EXAMPLE 2.6

The amount of oxalic acid in a sample of rhubarb was determined by reacting with Fe^{3+} as outlined in reaction 2.2. In a typical analysis, the oxalic acid in 10.62 g of rhubarb was extracted with a suitable solvent. The complete oxidation of the oxalic acid to CO_2 required 36.44 mL of 0.0130 M Fe^{3+}. What is the weight percent of oxalic acid in the sample of rhubarb?

SOLUTION

We begin by calculating the moles of Fe^{3+} used in the reaction

$$\frac{0.0130 \text{ mol } Fe^{3+}}{\text{L}} \times 0.03644 \text{ L} = 4.737 \times 10^{-4} \text{ mol } Fe^{3+}$$

The moles of oxalic acid reacting with the Fe^{3+}, therefore, is

$$4.737 \times 10^{-4} \text{ mol } Fe^{3+} \times \frac{1 \text{ mol } C_2H_2O_4}{2 \text{ mol } Fe^{3+}} = 2.369 \times 10^{-4} \text{ mol } C_2H_2O_4$$

Converting moles of oxalic acid to grams of oxalic acid

$$2.369 \times 10^{-4} \text{ mol } C_2H_2O_4 \times \frac{90.03 \text{ g } C_2H_2O_4}{\text{mol } C_2H_2O_4} = 2.132 \times 10^{-2} \text{ g oxalic acid}$$

and converting to weight percent gives the concentration of oxalic acid in the sample of rhubarb as

$$\frac{2.132 \times 10^{-2} \text{ g } C_2H_2O_4}{10.62 \text{ g rhubarb}} \times 100 = 0.201\% \text{ w/w } C_2H_2O_4$$

In the analysis described in Example 2.6 oxalic acid already was present in the desired form. In many analytical methods the compound to be determined must be converted to another form prior to analysis. For example, one method for the quantitative analysis of tetraethylthiuram disulfide ($C_{10}H_{20}N_2S_4$), the active ingredient in the drug Antabuse (disulfiram), requires oxidizing the S to SO_2, bubbling the SO_2 through H_2O_2 to produce H_2SO_4, followed by an acid–base titration of the H_2SO_4 with NaOH. Although we can write and balance chemical reactions for each of these steps, it often is easier to apply the principle of the conservation of reaction units.

A reaction unit is that part of a chemical species involved in a reaction. Consider, for example, the general unbalanced chemical reaction

$$A + B \rightarrow \text{Products}$$

Conservation of reaction units requires that the number of reaction units associated with the reactant A equal the number of reaction units associated with the reactant B. Translating the previous statement into mathematical form gives

$$\begin{aligned}&\text{Number of reaction units per A} \times \text{moles A} \\ &= \text{number of reaction units per B} \times \text{moles B}\end{aligned} \qquad \textbf{2.3}$$

If we know the moles of A and the number of reaction units associated with A and B, then we can calculate the moles of B. Note that a conservation of reaction units, as defined by equation 2.3, can only be applied between two species. There are five important principles involving a conservation of reaction units: mass, charge, protons, electron pairs, and electrons.

2C.1 Conservation of Mass

The easiest principle to appreciate is conservation of mass. Except for nuclear reactions, an element's total mass at the end of a reaction must be the same as that present at the beginning of the reaction; thus, an element serves as the most fundamental reaction unit. Consider, for example, the combustion of butane to produce CO_2 and H_2O, for which the unbalanced reaction is

$$C_4H_{10}(g) + O_2(g) \rightarrow CO_2(g) + H_2O(g)$$

All the carbon in CO_2 comes from the butane, thus we can select carbon as a reaction unit. Since there are four carbon atoms in butane, and one carbon atom in CO_2, we write

$$4 \times \text{moles } C_4H_{10} = 1 \times \text{moles } CO_2$$

Hydrogen also can be selected as a reaction unit since all the hydrogen in butane ends up in the H_2O produced during combustion. Thus, we can write

$$10 \times \text{moles } C_4H_{10} = 2 \times \text{moles } H_2O$$

Although the mass of oxygen is conserved during the reaction, we cannot apply equation 2.3 because the O_2 used during combustion does not end up in a single product.

Conservation of mass also can, with care, be applied to groups of atoms. For example, the ammonium ion, NH_4^+, can be precipitated as $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$. Selecting NH_4^+ as the reaction unit gives

$$2 \times \text{moles } Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O = 1 \times \text{moles } NH_4^+$$

2C.2 Conservation of Charge

The stoichiometry between two reactants in a precipitation reaction is governed by a conservation of charge, requiring that the total cation charge and the total anion charge in the precipitate be equal. The reaction units in a precipitation reaction, therefore, are the absolute values of the charges on the cation and anion that make up the precipitate. Applying equation 2.3 to a precipitate of $Ca_3(PO_4)_2$ formed from the reaction of Ca^{2+} and PO_4^{3-}, we write

$$2 \times \text{moles } Ca^{2+} = 3 \times \text{moles } PO_4^{3-}$$

2C.3 Conservation of Protons

In an acid–base reaction, the reaction unit is the proton. For an acid, the number of reaction units is given by the number of protons that can be donated to the base; and for a base, the number of reaction units is the number of protons that the base can accept from the acid. In the reaction between H_3PO_4 and NaOH, for example, the weak acid H_3PO_4 can donate all three of its protons to NaOH, whereas the strong base NaOH can accept one proton. Thus, we write

$$3 \times \text{moles } H_3PO_4 = 1 \times \text{moles NaOH}$$

Care must be exercised in determining the number of reaction units associated with the acid and base. The number of reaction units for an acid, for instance, depends not on how many acidic protons are present, but on how many

of the protons are capable of reacting with the chosen base. In the reaction between H_3PO_4 and NH_3

$$H_3PO_4(aq) + 2NH_3(aq) \rightleftharpoons HPO_4^-(aq) + 2NH_4^+(aq)$$

a conservation of protons requires that

$$2 \times \text{moles } H_3PO_4 = \text{moles of } NH_3$$

2C.4 Conservation of Electron Pairs

In a complexation reaction, the reaction unit is an electron pair. For the metal, the number of reaction units is the number of coordination sites available for binding ligands. For the ligand, the number of reaction units is equivalent to the number of electron pairs that can be donated to the metal. One of the most important analytical complexation reactions is that between the ligand ethylenediaminetetracetic acid (EDTA), which can donate 6 electron pairs and 6 coordinate metal ions, such as Cu^{2+}; thus

$$6 \times \text{mole } Cu^{2+} = 6 \times \text{moles EDTA}$$

2C.5 Conservation of Electrons

In a redox reaction, the reaction unit is an electron transferred from a reducing agent to an oxidizing agent. The number of reaction units for a reducing agent is equal to the number of electrons released during its oxidation. For an oxidizing agent, the number of reaction units is given by the number of electrons needed to cause its reduction. In the reaction between Fe^{3+} and oxalic acid (reaction 2.2), for example, Fe^{3+} undergoes a 1-electron reduction. Each carbon atom in oxalic acid is initially present in a +3 oxidation state, whereas the carbon atom in CO_2 is in a +4 oxidation state. Thus, we can write

$$1 \times \text{moles } Fe^{3+} = 2 \times \text{moles of } H_2C_2O_4$$

Note that the moles of oxalic acid are multiplied by 2 since there are two carbon atoms, each of which undergoes a 1-electron oxidation.

2C.6 Using Conservation Principles in Stoichiometry Problems

As shown in the following examples, the application of conservation principles simplifies stoichiometric calculations.

Example

EXAMPLE 2.7

Rework Example 2.6 using conservation principles.

SOLUTION

Conservation of electrons for this redox reaction requires that

$$\text{moles } Fe^{3+} = 2 \times \text{moles } H_2C_2O_4$$

which can be transformed by writing moles as the product of molarity and volume or as grams per formula weight.

$$M_{Fe^{3+}} \times V_{Fe^{3+}} = \frac{2 \times \text{g } H_2C_2O_4}{\text{FW } H_2C_2O_4}$$

Solving for g $H_2C_2O_4$ gives

$$\frac{M_{Fe^{3+}} \times V_{Fe^{3+}} \times FW\ H_2C_2O_4}{2} = \frac{(0.0130\ M)(0.03644\ L)(90.03\ g/mole)}{2}$$
$$= 2.132 \times 10^{-2}\ g\ H_2C_2O_4$$

and the weight percent oxalic acid is

$$\frac{2.132 \times 10^{-2}\ g\ C_2H_2O_4}{10.62\ g\ rhubarb} \times 100 = 0.201\%\ w/w\ C_2H_2O_4$$

Example

EXAMPLE 2.8

One quantitative analytical method for tetraethylthiuram disulfide, $C_{10}H_{20}N_2S_4$ (Antabuse), requires oxidizing the sulfur to SO_2, and bubbling the resulting SO_2 through H_2O_2 to produce H_2SO_4. The H_2SO_4 is then reacted with NaOH according to the reaction

$$H_2SO_4(aq) + 2NaOH(aq) \rightarrow Na_2SO_4(aq) + 2H_2O(\ell)$$

Using appropriate conservation principles, derive an equation relating the moles of $C_{10}H_{20}N_2S_4$ to the moles of NaOH. What is the weight percent $C_{10}H_{20}N_2S_4$ in a sample of Antabuse if the H_2SO_4 produced from a 0.4613-g portion reacts with 34.85 mL of 0.02500 M NaOH?

SOLUTION

The unbalanced reactions converting $C_{10}H_{20}N_2S_4$ to H_2SO_4 are

$$C_{10}H_{20}N_2S_4 \rightarrow SO_2$$
$$SO_2 \rightarrow H_2SO_4$$

Using a conservation of mass we have

$$4 \times \text{moles } C_{10}H_{20}N_2S_4 = \text{moles } SO_2 = \text{moles } H_2SO_4$$

A conservation of protons for the reaction of H_2SO_4 with NaOH gives

$$2 \times \text{moles } H_2SO_4 = \text{moles of NaOH}$$

Combining the two conservation equations gives the following stoichiometric equation between $C_{10}H_{20}N_2S_4$ and NaOH

$$8 \times \text{moles } C_{10}H_{20}N_2S_4 = \text{moles NaOH}$$

Now we are ready to finish the problem. Making appropriate substitutions for moles of $C_{10}H_{20}N_2S_4$ and moles of NaOH gives

$$\frac{8 \times g\ C_{10}H_{20}N_2S_4}{FW\ C_{10}H_{20}N_2S_4} = M_{NaOH} \times V_{NaOH}$$

Solving for g $C_{10}H_{20}N_2S_4$ gives

$$g\ C_{10}H_{20}N_2S_4 = \frac{1}{8} \times M_{NaOH} \times V_{NaOH} \times FW\ C_{10}H_{20}N_2S_4$$

$$\frac{1}{8}(0.02500\ M)(0.03485\ L)(296.54\ g/mol) = 0.032295\ g\ C_{10}H_{20}N_2S_4$$

The weight percent $C_{10}H_{20}N_2S_4$ in the sample, therefore, is

$$\frac{0.32295 \text{ g } C_{10}H_{20}N_2S_4}{0.4613 \text{ g sample}} \times 100 = 7.001\% \text{ w/w } C_{10}H_{20}N_2S_4$$

2D Basic Equipment and Instrumentation

Measurements are made using appropriate equipment or instruments. The array of equipment and instrumentation used in analytical chemistry is impressive, ranging from the simple and inexpensive, to the complex and costly. With two exceptions, we will postpone the discussion of equipment and instrumentation to those chapters where they are used. The instrumentation used to measure mass and much of the equipment used to measure volume are important to all analytical techniques and are therefore discussed in this section.

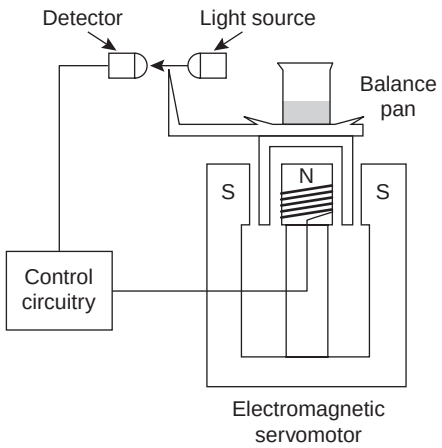
2D.1 Instrumentation for Measuring Mass

An object's mass is measured using a **balance.** The most common type of balance is an electronic balance in which the balance pan is placed over an electromagnet (Figure 2.2). The sample to be weighed is placed on the sample pan, displacing the pan downward by a force equal to the product of the sample's mass and the acceleration due to gravity. The balance detects this downward movement and generates a counterbalancing force using an electromagnet. The current needed to produce this force is proportional to the object's mass. A typical electronic balance has a capacity of 100–200 g and can measure mass to the nearest ±0.01 to ±1 mg.

balance
An apparatus used to measure mass.

Another type of balance is the single-pan, unequal arm balance (Figure 2.3). In this mechanical balance the balance pan and a set of removable standard weights on one side of a beam are balanced against a fixed counterweight on the beam's other side. The beam itself is balanced on a fulcrum consisting of a sharp knife edge. Adding a sample to the balance pan tilts the beam away from its balance point. Selected standard weights are then removed until the beam is brought back into balance. The combined mass of the removed weights equals the sample's mass. The capacities and measurement limits of these balances are comparable to an electronic balance.

(a)



(b)

Figure 2.2
(a) Photo of a typical electronic balance. (b) Schematic diagram of electronic balance; adding a sample moves the balance pan down, allowing more light to reach the detector. The control circuitry directs the electromagnetic servomotor to generate an opposing force, raising the sample up until the original intensity of light at the detector is restored.
Photo courtesy of Fisher Scientific.

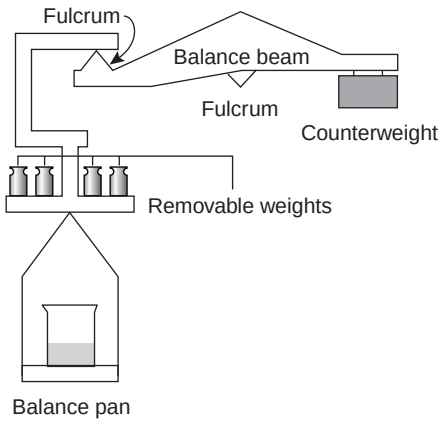


Figure 2.3
Schematic diagram of single-arm mechanical balance.

The mass of a sample is determined by difference. If the material being weighed is not moisture-sensitive, a clean and dry container is placed on the balance. The mass of this container is called the tare. Most balances allow the tare to be automatically adjusted to read a mass of zero. The sample is then transferred to the container, the new mass is measured and the sample's mass determined by subtracting the tare. Samples that absorb moisture from the air are weighed differently. The sample is placed in a covered weighing bottle and their combined mass is determined. A portion of the sample is removed, and the weighing bottle and remaining sample are reweighed. The difference between the two masses gives the mass of the transferred sample.

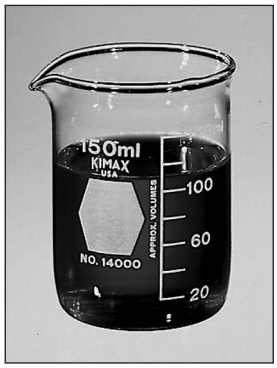
Several important precautions help to minimize errors in measuring an object's mass. Balances should be placed on heavy surfaces to minimize the effect of vibrations in the surrounding environment and should be maintained in a level position. Analytical balances are sensitive enough that they can measure the mass of a fingerprint. For this reason, materials placed on a balance should normally be handled using tongs or laboratory tissues. Volatile liquid samples should be weighed in a covered container to avoid the loss of sample by evaporation. Air currents can significantly affect a sample's mass. To avoid air currents, the balance's glass doors should be closed, or the balance's wind shield should be in place. A sample that is cooler or warmer than the surrounding air will create convective air currents that adversely affect the measurement of its mass. Finally, samples dried in an oven should be stored in a desiccator to prevent them from reabsorbing moisture from the atmosphere.

2D.2 Equipment for Measuring Volume

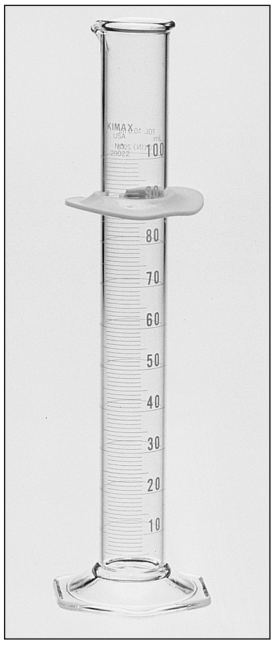
Analytical chemists use a variety of glassware to measure volume, several examples of which are shown in Figure 2.4. The type of glassware used depends on how exact the volume needs to be. Beakers, dropping pipets, and graduated cylinders are used to measure volumes approximately, typically with errors of several percent.

volumetric flask
Glassware designed to contain a specific volume of solution when filled to its calibration mark.

Pipets and volumetric flasks provide a more accurate means for measuring volume. When filled to its calibration mark, a **volumetric flask** is designed to contain a specified volume of solution at a stated temperature, usually 20 °C. The actual vol-

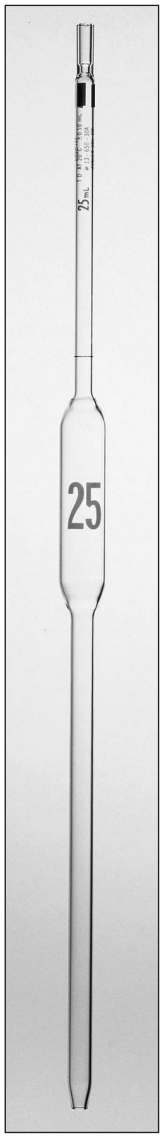


(a)

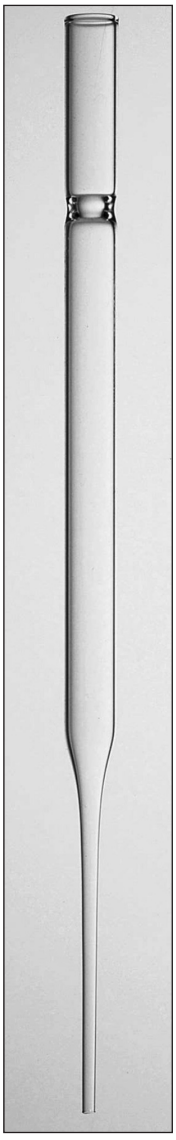


(b)

(c)



(d)



(e)

Figure 2.4
Common examples of glassware used to measure volume: (a) beaker; (b) graduated cylinder; (c) volumetric flask; (d) pipet; (e) dropping pipet.
Photos courtesy of Fisher Scientific.

ume contained by the volumetric flask is usually within 0.03–0.2% of the stated value. Volumetric flasks containing less than 100 mL generally measure volumes to the hundredth of a milliliter, whereas larger volumetric flasks measure volumes to the tenth of a milliliter. For example, a 10-mL volumetric flask contains 10.00 mL, but a 250-mL volumetric flask holds 250.0 mL (this is important when keeping track of significant figures).

Because a volumetric flask contains a solution, it is useful in preparing solutions with exact concentrations. The reagent is transferred to the volumetric flask, and enough solvent is added to dissolve the reagent. After the reagent is dissolved, additional solvent is added in several portions, mixing the solution after each addition. The final adjustment of volume to the flask's calibration mark is made using a dropping pipet. To complete the mixing process, the volumetric flask should be inverted at least ten times.

A **pipet** is used to deliver a specified volume of solution. Several different styles of pipets are available (Figure 2.5). Transfer pipets provide the most accurate means for delivering a known volume of solution; their volume error is similar to that from an equivalent volumetric flask. A 250-mL transfer pipet, for instance, will deliver 250.0 mL. To fill a transfer pipet, suction from a rubber bulb is used to pull the liquid up past the calibration mark (*never* use your mouth to suck a solution into a pipet). After replacing the bulb with your finger, the liquid's level is adjusted to the calibration mark, and the outside of the pipet is wiped dry. The pipet's contents are allowed to drain into the receiving container with the tip of the pipet touching the container walls. A small portion of the liquid remains in the pipet's tip and should not be blown out. Measuring pipets are used to deliver variable volumes, but with less accuracy than transfer pipets. With some measuring

pipet
Glassware designed to deliver a specific volume of solution when filled to its calibration mark.

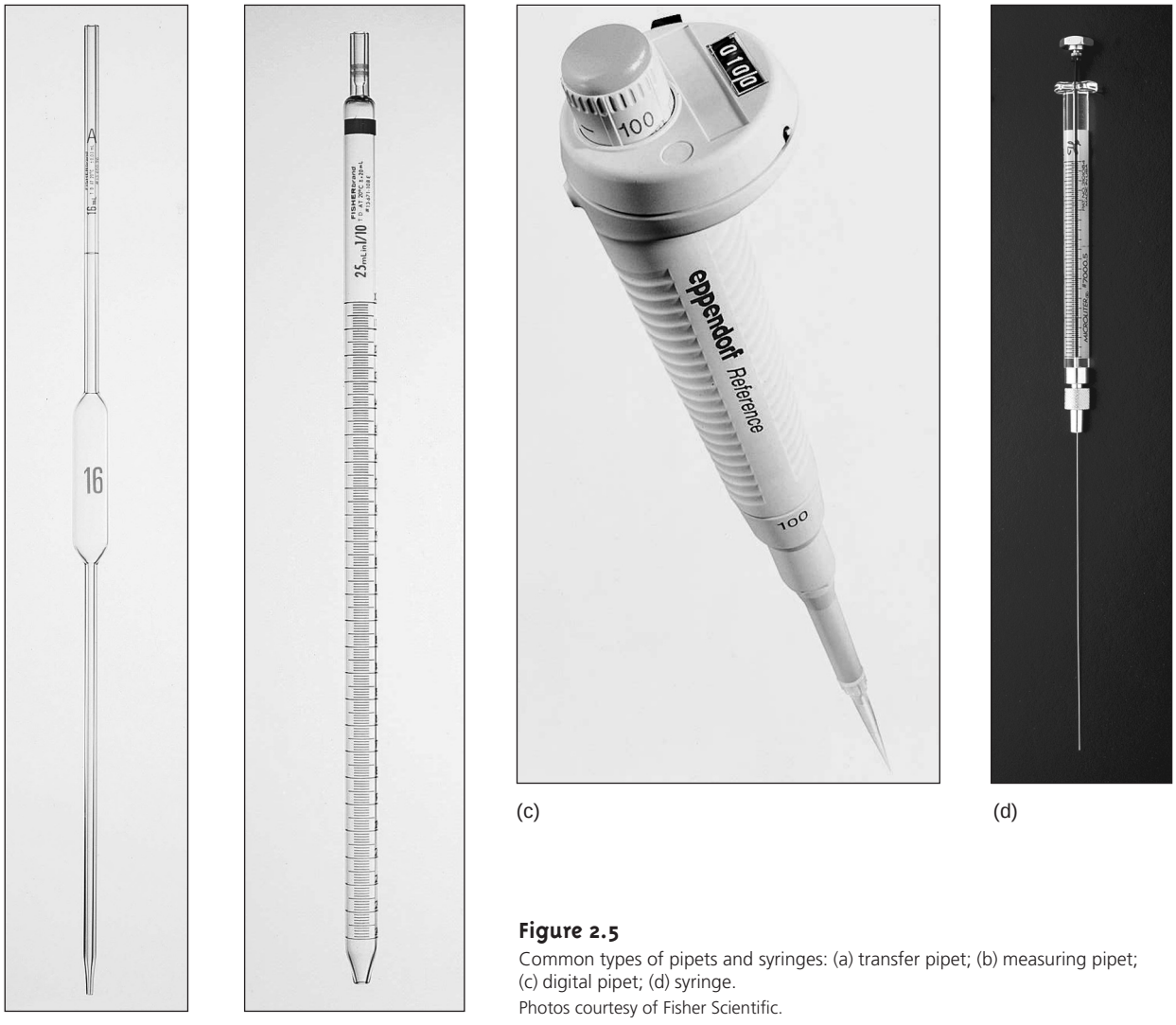


Figure 2.5
Common types of pipets and syringes: (a) transfer pipet; (b) measuring pipet; (c) digital pipet; (d) syringe.
Photos courtesy of Fisher Scientific.

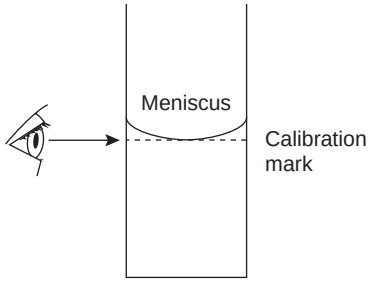


Figure 2.6
Proper means of reading the meniscus on a volumetric flask or pipet.

pipets, delivery of the calibrated volume requires that any solution remaining in the tip be blown out. Digital pipets and syringes can be used to deliver volumes as small as a microliter.

Three important precautions are needed when working with pipets and volumetric flasks. First, the volume delivered by a pipet or contained by a volumetric flask assumes that the glassware is clean. Dirt and grease on the inner glass surface prevents liquids from draining evenly, leaving droplets of the liquid on the container's walls. For a pipet this means that the delivered volume is less than the calibrated volume, whereas drops of liquid above the calibration mark mean that a volumetric flask contains more than its calibrated volume. Commercially available cleaning solutions can be used to clean pipets and volumetric flasks.

Second, when filling a pipet or volumetric flask, set the liquid's level exactly at the calibration mark. The liquid's top surface is curved into a **meniscus,** the bottom of which should be exactly even with the glassware's calibration mark (Figure 2.6). The meniscus should be adjusted with the calibration mark at eye level to avoid parallax errors. If your eye level is above the calibration mark the pipet or volumetric flask will be overfilled. The pipet or volumetric flask will be underfilled if your eye level is below the calibration mark.

Finally, before using a pipet or volumetric flask you should rinse it with several small portions of the solution whose volume is being measured. This ensures that any residual liquid remaining in the pipet or volumetric flask is removed.

meniscus
The curved surface of a liquid contained in a tube.

2D.3 Equipment for Drying Samples

Many materials need to be dried prior to their analysis to remove residual moisture. Depending on the material, heating to a temperature of 110–140 °C is usually sufficient. Other materials need to be heated to much higher temperatures to initiate thermal decomposition. Both processes can be accomplished using a laboratory oven capable of providing the required temperature.

Commercial laboratory ovens (Figure 2.7) are used when the maximum desired temperature is 160–325 °C (depending on the model). Some ovens include the ability to circulate heated air, allowing for a more efficient removal of moisture and shorter drying times. Other ovens provide a tight seal for the door, allowing the oven to be evacuated. In some situations a conventional laboratory oven can be replaced with a microwave oven. Higher temperatures, up to 1700° C, can be achieved using a muffle furnace (Figure 2.8).

Figure 2.7
Conventional laboratory oven used for drying materials.

After drying or decomposing a sample, it should be cooled to room temperature in a desiccator to avoid the readsorption of moisture. A **desiccator** (Figure 2.9) is a closed container that isolates the sample from the atmosphere. A drying agent, called a **desiccant,** is placed in the bottom of the container. Typical desiccants include calcium chloride and silica gel. A perforated plate sits above the desiccant, providing a shelf for storing samples. Some desiccators are equipped with stopcocks that allow them to be evacuated.

Figure 2.8
Example of a muffle furnace used for heating samples to maximum temperatures of 1100–1700 °C.
Courtesy of Fisher Scientific.

desiccator
A closed container containing a desiccant; used to store samples in a moisture-free environment.

desiccant
A drying agent.

(a) (b)

Figure 2.9
(a) Desiccator. (b) Desiccator with stopcock for evacuating the desiccator.
Photos courtesy of Fisher Scientific.

2E Preparing Solutions

Preparing a solution of known concentration is perhaps the most common activity in any analytical lab. The method for measuring out the solute and solvent depend on the desired concentration units, and how exact the solution's concentration needs to be known. Pipets and volumetric flasks are used when a solution's concentration must be exact; graduated cylinders, beakers, and reagent bottles suffice when concentrations need only be approximate. Two methods for preparing solutions are described in this section.

2E.1 Preparing Stock Solutions

stock solution
A solution of known concentration from which other solutions are prepared.

A **stock solution** is prepared by weighing out an appropriate portion of a pure solid or by measuring out an appropriate volume of a pure liquid and diluting to a known volume. Exactly how this is done depends on the required concentration units. For example, to prepare a solution with a desired molarity you would weigh out an appropriate mass of the reagent, dissolve it in a portion of solvent, and bring to the desired volume. To prepare a solution where the solute's concentration is given as a volume percent, you would measure out an appropriate volume of solute and add sufficient solvent to obtain the desired total volume.

Example

EXAMPLE 2.9

Describe how you would prepare the following three solutions: (a) 500 mL of approximately 0.20 M NaOH using solid NaOH; (b) 1 L of 150.0 ppm Cu^{2+} using Cu metal; and (c) 2 L of 4% v/v acetic acid using concentrated glacial acetic acid.

SOLUTION

(a) Since the concentration only needs to be known to two significant figures, the mass of NaOH and volume of solution do not need to be measured exactly. The desired mass of NaOH is

$$\frac{0.20 \text{ mol}}{\text{L}} \times \frac{40.0 \text{ g}}{\text{mol}} \times 0.50 \text{ L} = 4.0 \text{ g}$$

To prepare the solution we place 4.0 g of NaOH, weighed to the nearest tenth of a gram, in a bottle or beaker and add approximately 500 mL of water.

(b) Since the concentration of Cu^{2+} needs to be exact, the mass of Cu metal and the final solution volume must be measured exactly. The desired mass of Cu metal is

$$\frac{150.0 \text{ mg}}{\text{L}} \times 1.000 \text{ L} = 150.0 \text{ mg} = 0.1500 \text{ g}$$

To prepare the solution we measure out exactly 0.1500 g of Cu into a small beaker. To dissolve the Cu we add a small portion of concentrated HNO_3 and gently heat until it completely dissolves. The resulting solution is poured into a 1-L volumetric flask. The beaker is rinsed repeatedly with small portions of water, which are added to the volumetric flask. This process, which is called a **quantitative transfer,** ensures that the Cu^{2+} is completely transferred to the volumetric flask. Finally, additional water is added to the volumetric flask's calibration mark.

quantitative transfer
The process of moving a sample from one container to another in a manner that ensures all material is transferred.

(c) The concentration of this solution is only approximate, so volumes do not need to be measured exactly. The necessary volume of glacial acetic acid is

$$\frac{4 \text{ mL } CH_3COOH}{100 \text{ mL}} \times 2000 \text{ mL} = 80 \text{ mL } CH_3COOH$$

To prepare the solution we use a graduated cylinder to transfer 80 mL of glacial acetic acid to a container that holds approximately 2 L, and we then add sufficient water to bring the solution to the desired volume.

2E.2 Preparing Solutions by Dilution

Solutions with small concentrations are often prepared by diluting a more concentrated stock solution. A known volume of the stock solution is transferred to a new container and brought to a new volume. Since the total amount of solute is the same before and after **dilution,** we know that

dilution
The process of preparing a less concentrated solution from a more concentrated solution.

$$C_o \times V_o = C_d \times V_d \qquad \textbf{2.4}$$

where C_o is the concentration of the stock solution, V_o is the volume of the stock solution being diluted, C_d is the concentration of the dilute solution, and V_d is the volume of the dilute solution. Again, the type of glassware used to measure V_o and V_d depends on how exact the solution's concentration must be known.

Example

EXAMPLE 2.10

A laboratory procedure calls for 250 mL of an approximately 0.10 M solution of NH_3. Describe how you would prepare this solution using a stock solution of concentrated NH_3 (14.8 M).

SOLUTION

Substituting known volumes in equation 2.4

$$14.8 \text{ M} \times V_o = 0.10 \text{ M} \times 0.25 \text{ L}$$

and solving for V_o gives 1.69×10^{-3} L, or 1.7 mL. Since we are trying to make a solution that is approximately 0.10 M NH_3, we can measure the appropriate amount of concentrated NH_3 using a graduated cylinder, transfer the NH_3 to a beaker, and add sufficient water to bring the total solution volume to approximately 250 mL.

As shown in the following example, equation 2.4 also can be used to calculate a solution's original concentration using its known concentration after dilution.

Example

EXAMPLE 2.11

A sample of an ore was analyzed for Cu^{2+} as follows. A 1.25-g sample of the ore was dissolved in acid and diluted to volume in a 250-mL volumetric flask. A 20-mL portion of the resulting solution was transferred by pipet to a 50-mL volumetric flask and diluted to volume. An analysis showed that the concentration of Cu^{2+} in the final solution was 4.62 ppm. What is the weight percent of Cu in the original ore?

SOLUTION

Substituting known volumes (with significant figures appropriate for pipets and volumetric flasks) into equation 2.4

$$(\text{ppm } Cu^{2+})_o \times 20.00 \text{ mL} = 4.62 \text{ ppm} \times 50.00 \text{ mL}$$

and solving for $(\text{ppm } Cu^{2+})_o$ gives the original solution concentration as 11.55 ppm. To calculate the grams of Cu^{2+} we multiply this concentration by the total volume

$$\frac{11.55 \text{ μg } Cu^{2+}}{\text{mL}} \times 250.0 \text{ mL} \times \frac{1 \text{ g}}{10^6 \text{ μg}} = 2.888 \times 10^{-3} \text{ g } Cu^{2+}$$

The weight percent Cu is then given by

$$\frac{2.888 \times 10^{-3} \text{ g } Cu^{2+}}{1.25 \text{ g sample}} \times 100 = 0.231\% \text{ w/w Cu}$$

2F The Laboratory Notebook

Finally, we cannot end a chapter on the basic tools of analytical chemistry without mentioning the laboratory notebook. Your laboratory notebook is your most important tool when working in the lab, providing a complete record of all your work. If kept properly, you should be able to look back at your laboratory notebook several years from now and reconstruct the experiments on which you worked.

Your instructor will probably provide you with detailed instructions on how he or she wants you to maintain your notebook. Of course, you should expect to bring your notebook to the lab. Everything you do, measure, or observe while working in the lab should be recorded in your notebook as it takes place. Preparing data tables to organize your data will help ensure that you record the data you need and that you can find the data when it is time to calculate and analyze your results. Writing a narrative to accompany your data will help you remember what you did, why you did it, and why you thought it was significant. Reserve space for your calculations, for analyzing your data, and for interpreting your results. Take your notebook with you when you do research in the library.

Maintaining a laboratory notebook may seem like a great deal of effort, but if you do it well you have a permanent record of your work. Scientists working in academic, industrial, and governmental research labs rely on their notebooks to provide a written record of their work. Questions about research carried out at some time in the past can be answered by finding the appropriate pages in the laboratory notebook. A laboratory notebook is also a legal document that helps establish patent rights and proof of discovery.

2G KEY TERMS

balance (*p. 25*)
concentration (*p. 15*)
desiccant (*p. 29*)
desiccator (*p. 29*)
dilution (*p. 31*)
equivalent (*p. 17*)
equivalent weight (*p. 17*)
formality (*p. 15*)
formula weight (*p. 17*)
meniscus (*p. 29*)
molality (*p. 18*)
molarity (*p. 15*)

normality (*p. 16*)
parts per billion (*p. 18*)
parts per million (*p. 18*)
p-function (*p. 19*)
pipet (*p. 27*)
quantitative transfer (*p. 30*)
scientific notation (*p. 12*)
significant figures (*p. 13*)
SI units (*p. 12*)
stock solution (*p. 30*)
volume percent (*p. 18*)
volumetric flask (*p. 26*)
weight percent (*p. 18*)
weight-to-volume percent (*p. 18*)

2H SUMMARY

There are a few basic numerical and experimental tools with which you must be familiar. Fundamental measurements in analytical chemistry, such as mass and volume, use base SI units, such as the kilogram (kg) and the liter (L). Other units, such as power, are defined in terms of these base units. When reporting measurements, we must be careful to include only those digits that are significant and to maintain the uncertainty implied by these significant figures when transforming measurements into results.

The relative amount of a constituent in a sample is expressed as its concentration. There are many ways to express concentration, the most common of which are molarity, weight percent, volume percent, weight-to-volume percent, parts per million, and parts per billion. Concentrations also can be expressed using p-functions.

Stoichiometric relationships and calculations are important in many quantitative analyses. The stoichiometry between the reactants and products of a chemical reaction is given by the coefficients of a balanced chemical reaction. When it is inconvenient to balance reactions, conservation principles can be used to establish the stoichiometric relationships.

Balances, volumetric flasks, pipets, and ovens are standard pieces of laboratory instrumentation and equipment that are routinely used in almost all analytical work. You should be familiar with the proper use of this equipment. You also should be familiar with how to prepare a stock solution of known concentration, and how to prepare a dilute solution from a stock solution.

2I PROBLEMS

1. Indicate how many significant figures are in each of the following numbers.
   a. 903 b. 0.903 c. 1.0903
   d. 0.0903 e. 0.09030 f. 9.03×10^2

2. Round each of the following to three significant figures.
   a. 0.89377 b. 0.89328 c. 0.89350
   d. 0.8997 e. 0.08907

3. Round each of the following to the stated number of significant figures.
   a. The atomic weight of carbon to four significant figures
   b. The atomic weight of oxygen to three significant figures
   c. Avogadro's number to four significant figures
   d. Faraday's constant to three significant figures

4. Report results for the following calculations to the correct number of significant figures.
   a. $4.591 + 0.2309 + 67.1 =$
   b. $313 - 273.15 =$
   c. $712 \times 8.6 =$
   d. $1.43/0.026 =$
   e. $(8.314 \times 298)/96485 =$
   f. $\log(6.53 \times 10^{-5}) =$
   g. $10^{-7.14} =$
   h. $(6.51 \times 10^{-5})\,(8.14 \times 10^{-9}) =$

5. A 12.1374-g sample of an ore containing Ni and Co was carried through Fresenius' analytical scheme shown in Figure 1.1. At point A the combined mass of Ni and Co was found to be 0.2306 g, and at point B the mass of Co was found to be 0.0813 g. Report the weight percent Ni in the ore to the correct number of significant figures.

6. Hillebrand and Lundell's analytical scheme (see Figure 1.2) for the analysis of Ni in ores involves precipitating Ni^{2+} using dimethylgloxime. The formula for the precipitate is $Ni(C_4H_7N_2O_2)_2$. Calculate the precipitate's formula weight to the correct number of significant figures.

7. An analyst wishes to add 256 mg of Cl^- to a reaction mixture. How many milliliters of 0.217 M $BaCl_2$ should be added?

8. A solution of 0.10 M SO_4^{2-} is available. What is the normality of this solution when used in the following reactions?
   a. $Pb^{2+}(aq) + SO_4^{2-}(aq) \rightleftharpoons PbSO_4(s)$
   b. $HCl(aq) + SO_4^{2-}(aq) \rightleftharpoons HSO_4^-(aq) + Cl^-(aq)$
   c. $SO_4^{2-} + 4H_3O^+(aq) + 2e^- \rightleftharpoons H_2SO_3(aq) + 5H_2O(\ell)$

9. The concentration of lead in an industrial waste stream is 0.28 ppm. What is its molar concentration?

10. Commercially available concentrated hydrochloric acid is 37.0% w/w HCl. Its density is 1.18 g/mL. Using this information calculate (a) the molarity of concentrated HCl, and (b) the mass and volume (in milliliters) of solution containing 0.315 mol of HCl.

11. The density of concentrated ammonia, which is 28.0% w/w NH_3, is 0.899 g/mL. What volume of this reagent should be diluted to 1.0×10^3 mL to make a solution that is 0.036 M in NH_3?

12. A 250.0-mL aqueous solution contains 45.1 μg of a pesticide. Express the pesticide's concentration in weight percent, parts per million, and parts per billion.

13. A city's water supply is fluoridated by adding NaF. The desired concentration of F^- is 1.6 ppm. How many milligrams of NaF should be added per gallon of treated water if the water supply already is 0.2 ppm in F^-?

14. What is the pH of a solution for which the concentration of H^+ is 6.92×10^{-6} M? What is the $[H^+]$ in a solution whose pH is 8.923?

15. Using conservation principles, write stoichiometric relationships for the following
   a. The precipitation of Mg^{2+} as $Mg_2P_2O_7$
   b. The acid–base reaction between $CaCO_3$ and HCl in which H_2CO_3 is formed
   c. The reaction between AgCl and NH_3 to form $Ag(NH_3)_2^+$
   d. The redox reaction between $Cr_2O_7^{2-}$ and Fe^{2+} to form Cr^{3+} and Fe^{3+}

16. Calculate the molarity of a potassium dichromate solution prepared by placing 9.67 g of $K_2Cr_2O_7$ in a 100-mL volumetric flask, dissolving, and diluting to the calibration mark.

17. For each of the following, explain how you would prepare 1.0 L of a solution that is 0.10 M in K^+. Repeat for concentrations of 1.0×10^2 ppm K^+ and 1.0% w/v K^+.
   a. KCl b. K_2SO_4 c. $K_3Fe(CN)_6$

18. A series of dilute NaCl solutions is prepared, starting with an initial stock solution of 0.100 M NaCl. Solution A is prepared by pipeting 10 mL of the stock solution into a 250-mL volumetric flask and diluting to volume. Solution B is prepared by pipeting 25 mL of solution A into a 100-mL volumetric flask and diluting to volume. Solution C is prepared by pipeting 20 mL of solution B into a 500-mL volumetric flask and diluting to volume. What is the molar concentration of NaCl in solutions A, B, and C?

19. Calculate the molar concentration of NaCl, to the correct number of significant figures, if 1.917 g of NaCl is placed in a beaker and dissolved in 50 mL of water measured with a graduated cylinder. This solution is quantitatively transferred to a 250-mL volumetric flask and diluted to volume. Calculate the concentration of this second solution to the correct number of significant figures.

20. What is the molar concentration of NO_3^- in a solution prepared by mixing 50.0 mL of 0.050 M KNO_3 with 40.0 mL of 0.075 M $NaNO_3$? What is pNO_3 for the mixture?

21. What is the molar concentration of Cl^- in a solution prepared by mixing 25.0 mL of 0.025 M NaCl with 35.0 mL of 0.050 M $BaCl_2$? What is pCl for the mixture?

22. To determine the concentration of ethanol in cognac a 5.00-mL sample of cognac is diluted to 0.500 L. Analysis of the diluted cognac gives an ethanol concentration of 0.0844 M. What is the molar concentration of ethanol in the undiluted cognac?

2J SUGGESTED READINGS

Two useful articles providing additional information on topics covered in this chapter are

MacCarthy, P. "A Novel Classification of Concentration Units," *J. Chem. Educ.* **1983,** *60,* 187–189.

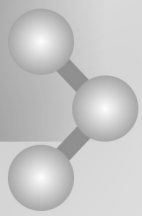
Schwartz, L. M. "Propagation of Significant Figures," *J. Chem. Educ.* **1985,** *62,* 693–697.

A useful resource for information on maintaining a useful laboratory notebook is

Kanare, H. M. *Writing the Laboratory Notebook,* American Chemical Society: Washington, DC; 1985.

2K REFERENCES

1. American Public Health Association. *Standard Methods for the Analysis of Waters and Wastewaters,* 19th ed., Washington, DC. 1995.



Chapter 3

The Language of Analytical Chemistry

Analytical chemists converse using terminology that conveys specific meaning to other analytical chemists. To discuss and learn analytical chemistry you must first understand its language. You are probably already familiar with some analytical terms, such as "accuracy" and "precision," but you may not have placed them in their appropriate analytical context. Other terms, such as "analyte" and "matrix," may be less familiar. This chapter introduces many important terms routinely used by analytical chemists. Becoming comfortable with these terms will make the material in the chapters that follow easier to read and understand.

3A Analysis, Determination, and Measurement

analysis
A process that provides chemical or physical information about the constituents in the sample or the sample itself.

analytes
The constituents of interest in a sample.

matrix
All other constituents in a sample except for the analytes.

determination
An analysis of a sample to find the identity, concentration, or properties of the analyte.

measurement
An experimental determination of an analyte's chemical or physical properties.

The first important distinction we will make is among the terms "analysis," "determination," and "measurement." An **analysis** provides chemical or physical information about a sample. The components of interest in the sample are called **analytes,** and the remainder of the sample is the **matrix.** In an analysis we determine the identity, concentration, or properties of the analytes. To make this **determination** we measure one or more of the analyte's chemical or physical properties.

An example helps clarify the differences among an analysis, a determination, and a **measurement.** In 1974, the federal government enacted the Safe Drinking Water Act to ensure the safety of public drinking water supplies. To comply with this act municipalities regularly monitor their drinking water supply for potentially harmful substances. One such substance is coliform bacteria. Municipal water departments collect and analyze samples from their water supply. To determine the concentration of coliform bacteria, a portion of water is passed through a membrane filter. The filter is placed in a dish containing a nutrient broth and incubated. At the end of the incubation period the number of coliform bacterial colonies in the dish is measured by counting (Figure 3.1). Thus, municipal water departments analyze samples of water to determine the concentration of coliform bacteria by measuring the number of bacterial colonies that form during a specified period of incubation.

3B Techniques, Methods, Procedures, and Protocols

Suppose you are asked to develop a way to determine the concentration of lead in drinking water. How would you approach this problem? To answer this question it helps to distinguish among four levels of analytical methodology: techniques, methods, procedures, and protocols.[1]

technique
A chemical or physical principle that can be used to analyze a sample.

A **technique** is any chemical or physical principle that can be used to study an analyte. Many techniques have been used to determine lead levels.[2] For example, in graphite furnace atomic absorption spectroscopy lead is atomized, and the ability of the free atoms to absorb light is measured; thus, both a chemical principle (atomization) and a physical principle (absorption of light) are used in this technique. Chapters 8–13 of this text cover techniques commonly used to analyze samples.

method
A means for analyzing a sample for a specific analyte in a specific matrix.

A **method** is the application of a technique for the determination of a specific analyte in a specific matrix. As shown in Figure 3.2, the graphite furnace atomic absorption spectroscopic method for determining lead levels in water is different from that for the determination of lead in soil or blood. Choosing a method for determining lead in water depends on how the information is to be used and the established design criteria (Figure 3.3). For some analytical problems the best method might use graphite furnace atomic absorption spectroscopy, whereas other problems might be more easily solved by using another technique, such as anodic stripping voltammetry or potentiometry with a lead ion-selective electrode.

procedure
Written directions outlining how to analyze a sample.

A **procedure** is a set of written directions detailing how to apply a method to a particular sample, including information on proper sampling, handling of interferents, and validating results. A method does not necessarily lead to a single procedure, as different analysts or agencies will adapt the method to their specific needs. As shown in Figure 3.2, the American Public Health Agency and the American Society for Testing Materials publish separate procedures for the determination of lead levels in water.

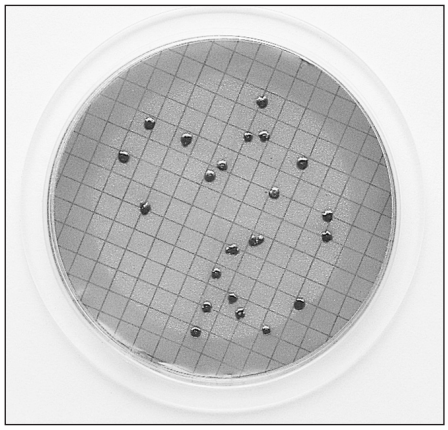


Figure 3.1
Membrane filter showing colonies of coliform bacteria. The number of colonies are counted and reported as colonies/100 mL of sample.
PourRite™ is a trademark of Hach Company/photo courtesy of Hach Company.

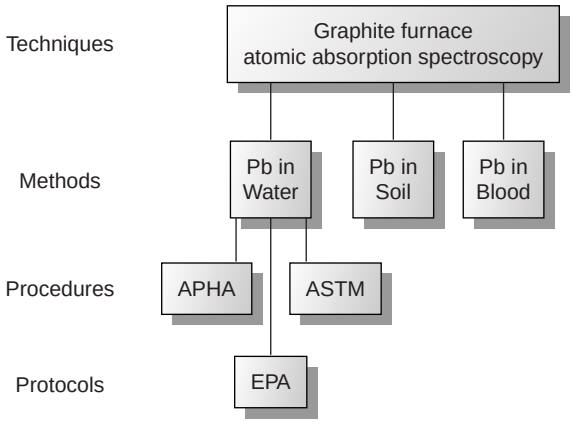


Figure 3.2
Chart showing hierarchical relationship among a technique, methods using that technique, and procedures and protocols for one method. (*Abbreviations:* APHA = American Public Health Association, ASTM = American Society for Testing Materials, EPA = Environmental Protection Agency)

Finally, a **protocol** is a set of stringent written guidelines detailing the procedure that must be followed if the agency specifying the protocol is to accept the results of the analysis. Protocols are commonly encountered when analytical chemistry is used to support or define public policy. For purposes of determining lead levels in water under the Safe Drinking Water Act, labs follow a protocol specified by the Environmental Protection Agency.

There is an obvious order to these four facets of analytical methodology. Ideally, a protocol uses a previously validated procedure. Before developing and validating a procedure, a method of analysis must be selected. This requires, in turn, an initial screening of available techniques to determine those that have the potential for monitoring the analyte. We begin by considering a useful way to classify analytical techniques.

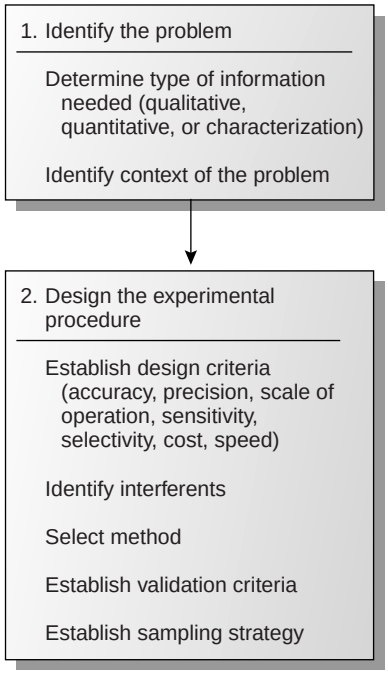


Figure 3.3
Subsection of the analytical approach to problem solving (see Figure 1.3), of relevance to the selection of a method and the design of an analytical procedure.

3C Classifying Analytical Techniques

Analyzing a sample generates a chemical or physical **signal** whose magnitude is proportional to the amount of analyte in the sample. The signal may be anything we can measure; common examples are mass, volume, and absorbance. For our purposes it is convenient to divide analytical techniques into two general classes based on whether this signal is proportional to an absolute amount of analyte or a relative amount of analyte.

Consider two graduated cylinders, each containing 0.01 M $Cu(NO_3)_2$ (Figure 3.4). Cylinder 1 contains 10 mL, or 0.0001 mol, of Cu^{2+}; cylinder 2 contains 20 mL, or 0.0002 mol, of Cu^{2+}. If a technique responds to the absolute amount of analyte in the sample, then the signal due to the analyte, S_A, can be expressed as

$$S_A = kn_A \qquad \textbf{3.1}$$

where n_A is the moles or grams of analyte in the sample, and k is a proportionality constant. Since cylinder 2 contains twice as many moles of Cu^{2+} as cylinder 1, analyzing the contents of cylinder 2 gives a signal that is twice that of cylinder 1.

protocol
A set of written guidelines for analyzing a sample specified by an agency.

signal
An experimental measurement that is proportional to the amount of analyte (S).

A second class of analytical techniques are those that respond to the relative amount of analyte; thus

$$S_A = kC_A \qquad \textbf{3.2}$$

where C_A is the concentration of analyte in the sample. Since the solutions in both cylinders have the same concentration of Cu^{2+}, their analysis yields identical signals.

total analysis techniques
A technique in which the signal is proportional to the absolute amount of analyte; also called "classical" techniques.

Techniques responding to the absolute amount of analyte are called **total analysis techniques.** Historically, most early analytical methods used total analysis techniques, hence they are often referred to as "classical" techniques. Mass, volume, and charge are the most common signals for total analysis techniques, and the corresponding techniques are gravimetry (Chapter 8), titrimetry (Chapter 9), and coulometry (Chapter 11). With a few exceptions, the signal in a total analysis technique results from one or more chemical reactions involving the analyte. These reactions may involve any combination of precipitation, acid–base, complexation, or redox chemistry. The stoichiometry of each reaction, however, must be known to solve equation 3.1 for the moles of analyte.

concentration techniques
A technique in which the signal is proportional to the analyte's concentration; also called "instrumental" techniques.

Techniques, such as spectroscopy (Chapter 10), potentiometry (Chapter 11), and voltammetry (Chapter 11), in which the signal is proportional to the relative amount of analyte in a sample are called **concentration techniques.** Since most concentration techniques rely on measuring an optical or electrical signal, they also are known as "instrumental" techniques. For a concentration technique, the relationship between the signal and the analyte is a theoretical function that depends on experimental conditions and the instrumentation used to measure the signal. For this reason the value of k in equation 3.2 must be determined experimentally.

(a) (b)

Figure 3.4
Graduated cylinders containing 0.01 M $Cu(NO_3)_2$. (a) Cylinder 1 contains 10 mL, or 0.0001 mol, of Cu^{2+}. (b) Cylinder 2 contains 20 mL, or 0.0002 mol, of Cu^{2+}.

3D Selecting an Analytical Method

A method is the application of a technique to a specific analyte in a specific matrix. Methods for determining the concentration of lead in drinking water can be developed using any of the techniques mentioned in the previous section. Insoluble lead salts such as $PbSO_4$ and $PbCrO_4$ can form the basis for a gravimetric method. Lead forms several soluble complexes that can be used in a complexation titrimetric method or, if the complexes are highly absorbing, in a spectrophotometric method. Lead in the gaseous free-atom state can be measured by an atomic absorption spectroscopic method. Finally, the availability of multiple oxidation states (Pb, Pb^{2+}, Pb^{4+}) makes coulometric, potentiometric, and voltammetric methods feasible.

The requirements of the analysis determine the best method. In choosing a method, consideration is given to some or all the following design criteria: accuracy, precision, sensitivity, selectivity, robustness, ruggedness, scale of operation, analysis time, availability of equipment, and cost. Each of these criteria is considered in more detail in the following sections.

3D.1 Accuracy

Accuracy is a measure of how closely the result of an experiment agrees with the expected result. The difference between the obtained result and the expected result is usually divided by the expected result and reported as a percent relative error

accuracy
A measure of the agreement between an experimental result and its expected value.

$$\%\ \text{Error} = \frac{\text{obtained result} - \text{expected result}}{\text{expected result}} \times 100$$

Analytical methods may be divided into three groups based on the magnitude of their relative errors.[3] When an experimental result is within 1% of the correct result, the analytical method is highly accurate. Methods resulting in relative errors between 1% and 5% are moderately accurate, but methods of low accuracy produce relative errors greater than 5%.

The magnitude of a method's relative error depends on how accurately the signal is measured, how accurately the value of k in equations 3.1 or 3.2 is known, and the ease of handling the sample without loss or contamination. In general, total analysis methods produce results of high accuracy, and concentration methods range from high to low accuracy. A more detailed discussion of accuracy is presented in Chapter 4.

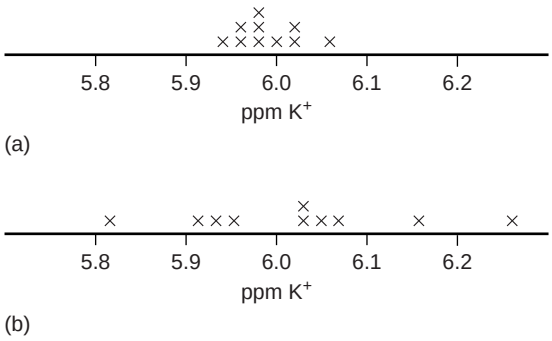


Figure 3.5
Two determinations of the concentration of K^+ in serum, showing the effect of precision. The data in (a) are less scattered and, therefore, more precise than the data in (b).

3D.2 Precision

When a sample is analyzed several times, the individual results are rarely the same. Instead, the results are randomly scattered. **Precision** is a measure of this variability. The closer the agreement between individual analyses, the more precise the results. For example, in determining the concentration of K^+ in serum, the results shown in Figure 3.5(a) are more precise than those in Figure 3.5(b). It is important to realize that precision does not imply accuracy. That the data in Figure 3.5(a) are more precise does not mean that the first set of results is more accurate. In fact, both sets of results may be very inaccurate.

precision
An indication of the reproducibility of a measurement or result.

As with accuracy, precision depends on those factors affecting the relationship between the signal and the analyte (equations 3.1 and 3.2). Of particular importance are the uncertainty in measuring the signal and the ease of handling samples reproducibly. In most cases the signal for a total analysis method can be measured with a higher precision than the corresponding signal for a concentration method. Precision is covered in more detail in Chapter 4.

3D.3 Sensitivity

The ability to demonstrate that two samples have different amounts of analyte is an essential part of many analyses. A method's **sensitivity** is a measure of its ability to establish that such differences are significant. Sensitivity is often confused with a method's detection limit.[4] The **detection limit** is the smallest amount of analyte that can be determined with confidence. The detection limit, therefore, is a statistical parameter and is discussed in Chapter 4.

sensitivity
A measure of a method's ability to distinguish between two samples; reported as the change in signal per unit change in the amount of analyte (k).

detection limit
A statistical statement about the smallest amount of analyte that can be determined with confidence.

Sensitivity is the change in signal per unit change in the amount of analyte and is equivalent to the proportionality constant, k, in equations 3.1 and 3.2. If ΔS_A is the smallest increment in signal that can be measured, then the smallest difference in the amount of analyte that can be detected is

$$\Delta n_A = \frac{\Delta S_A}{k} \quad \text{(total analysis method)}$$

$$\Delta C_A = \frac{\Delta S_A}{k} \quad \text{(concentration method)}$$

Suppose that for a particular total analysis method the signal is a measurement of mass using a balance whose smallest increment is ±0.0001 g. If the method's

sensitivity is 0.200, then the method can conceivably detect a difference of as little as

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

in the absolute amount of analyte in two samples. For methods with the same ΔS_A, the method with the greatest sensitivity is best able to discriminate among smaller amounts of analyte.

3D.4 Selectivity

An analytical method is selective if its signal is a function of only the amount of analyte present in the sample. In the presence of an interferent, equations 3.1 and 3.2 can be expanded to include a term corresponding to the interferent's contribution to the signal, S_I,

$$S_{samp} = S_A + S_I = k_A n_A + k_I n_I \qquad \text{(total analysis method)} \qquad \textbf{3.3}$$

$$S_{samp} = S_A + S_I = k_A C_A + k_I C_I \qquad \text{(concentration method)} \qquad \textbf{3.4}$$

where S_{samp} is the total signal due to constituents in the sample; k_A and k_I are the sensitivities for the analyte and the interferent, respectively; and n_I and C_I are the moles (or grams) and concentration of the interferent in the sample.

selectivity
A measure of a method's freedom from interferences as defined by the method's selectivity coefficient.

selectivity coefficient
A measure of a method's sensitivity for an interferent relative to that for the analyte ($K_{A,I}$).

The **selectivity** of the method for the interferent relative to the analyte is defined by a **selectivity coefficient,** $K_{A,I}$

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

which may be positive or negative depending on whether the interferent's effect on the signal is opposite that of the analyte.* A selectivity coefficient greater than +1 or less than –1 indicates that the method is more selective for the interferent than for the analyte. Solving equation 3.5 for k_I

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

substituting into equations 3.3 and 3.4, and simplifying gives

$$S_{samp} = k_A(n_A + K_{A,I} \times n_I) \qquad \text{(total analysis method)} \qquad \textbf{3.7}$$

$$S_{samp} = k_A(C_A + K_{A,I} \times C_I) \qquad \text{(concentration method)} \qquad \textbf{3.8}$$

The selectivity coefficient is easy to calculate if k_A and k_I can be independently determined. It is also possible to calculate $K_{A,I}$ by measuring S_{samp} in the presence and absence of known amounts of analyte and interferent.

Example

EXAMPLE 3.1

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

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

SOLUTION

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

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

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

In the presence of zinc the signal is

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

Solving for k_{Zn} gives a value of 0.5. The selectivity coefficient, therefore, is

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

Knowing the selectivity coefficient provides a useful way to evaluate an interferent's potential effect on an analysis. An interferent will not pose a problem as long as the term $K_{A,I} \times n_I$ in equation 3.7 is significantly smaller than n_A, or $K_{A,I} \times C_I$ in equation 3.8 is significantly smaller than C_A.

Example

EXAMPLE 3.2

Barnett and colleagues[5] developed a new method for determining the concentration of codeine during its extraction from poppy plants. As part of their study they determined the method's response to codeine relative to that for several potential interferents. For example, the authors found that the method's signal for 6-methoxycodeine was 6 (arbitrary units) when that for an equimolar solution of codeine was 40.

(a) What is the value for the selectivity coefficient $K_{A,I}$ when 6-methoxycodeine is the interferent and codeine is the analyte?

(b) If the concentration of codeine is to be determined with an accuracy of ±0.50%, what is the maximum relative concentration of 6-methoxycodeine (i.e., [6-methoxycodeine]/[codeine]) that can be present?

SOLUTION

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

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

Solving these two expressions for k_A and k_I and substituting into equation 3.6 gives

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

Since equimolar concentrations of analyte and interferent were used ($C_A = C_I$), we have

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

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

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

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

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

Therefore, the concentration of 6-methoxycodeine cannot exceed 3.3% of codeine's concentration.

Not surprisingly, methods whose signals depend on chemical reactivity are often less selective and, therefore, more susceptible to interferences. Problems with selectivity become even greater when the analyte is present at a very low concentration.[6]

3D.5 Robustness and Ruggedness

For a method to be useful it must provide reliable results. Unfortunately, methods are subject to a variety of chemical and physical interferences that contribute uncertainty to the analysis. When a method is relatively free from chemical interferences, it can be applied to the determination of analytes in a wide variety of sample matrices. Such methods are considered **robust.**

robust
A method that can be applied to analytes in a wide variety of matrices is considered robust.

Random variations in experimental conditions also introduce uncertainty. If a method's sensitivity is highly dependent on experimental conditions, such as temperature, acidity, or reaction time, then slight changes in those conditions may lead to significantly different results. A **rugged** method is relatively insensitive to changes in experimental conditions.

rugged
A method that is insensitive to changes in experimental conditions is considered rugged.

3D.6 Scale of Operation

Another way to narrow the choice of methods is to consider the scale on which the analysis must be conducted. Three limitations of particular importance are the amount of sample available for the analysis, the concentration of analyte in the sample, and the absolute amount of analyte needed to obtain a measurable signal. The first and second limitations define the scale of operations shown in Figure 3.6; the last limitation positions a method within the scale of operations.[7]

The scale of operations in Figure 3.6 shows the analyte's concentration in weight percent on the *y*-axis and the sample's size on the *x*-axis. For convenience, we divide analytes into major (>1% w/w), minor (0.01% w/w – 1% w/w), trace (10^{-7}% w/w – 0.01% w/w) and ultratrace (<10^{-7}% w/w) components, and we divide samples into macro (>0.1 g), meso (10 mg – 100 mg), micro (0.1 mg – 10 mg) and ultramicro (<0.1 mg) sample sizes. Note that both the *x*-axis and the *y*-axis use a logarithmic scale. The analyte's concentration and the amount of

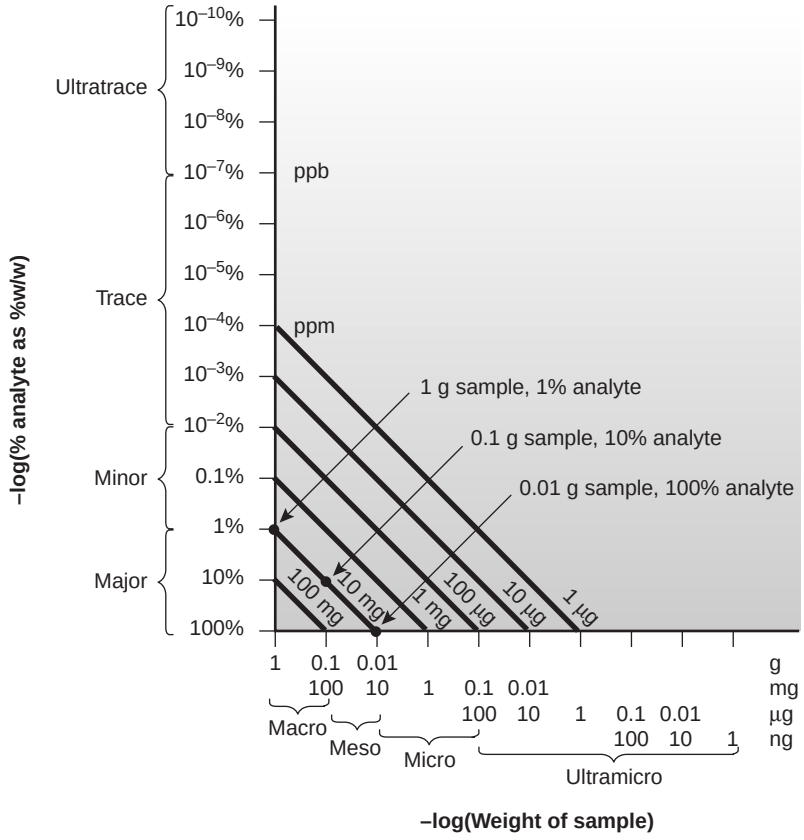


Figure 3.6
Scale of operation for analytical methods. Adapted from references 7*a* and 7*b*.

sample used provide a characteristic description for an analysis. For example, samples in a macro–major analysis weigh more than 0.1 g and contain more than 1% analyte.

Diagonal lines connecting the two axes show combinations of sample size and concentration of analyte containing the same absolute amount of analyte. As shown in Figure 3.6, for example, a 1-g sample containing 1% analyte has the same amount of analyte (0.010 g) as a 100-mg sample containing 10% analyte or a 10-mg sample containing 100% analyte.

Since total analysis methods respond to the absolute amount of analyte in a sample, the diagonal lines provide an easy way to define their limitations. Consider, for example, a hypothetical total analysis method for which the minimum detectable signal requires 100 mg of analyte. Using Figure 3.6, the diagonal line representing 100 mg suggests that this method is best suited for macro samples and major analytes. Applying the method to a minor analyte with a concentration of 0.1% w/w requires a sample of at least 100 g. Working with a sample of this size is rarely practical, however, due to the complications of carrying such a large amount of material through the analysis. Alternatively, the minimum amount of required analyte can be decreased by improving the limitations associated with measuring the signal. For example, if the signal is a measurement of mass, a decrease in the minimum amount of analyte can be accomplished by switching from a conventional analytical balance, which weighs samples to ±0.1 mg, to a semimicro (±0.01 mg) or microbalance (±0.001 mg).

Concentration methods frequently have both lower and upper limits for the amount of analyte that can be determined. The lower limit is dictated by the smallest concentration of analyte producing a useful signal and typically is in the parts per million or parts per billion concentration range. Upper concentration limits exist when the sensitivity of the analysis decreases at higher concentrations.

An upper concentration level is important because it determines how a sample with a high concentration of analyte must be treated before the analysis. Consider, for example, a method with an upper concentration limit of 1 ppm (micrograms per milliliter). If the method requires a sample of 1 mL, then the upper limit on the amount of analyte that can be handled is 1 μg. Using Figure 3.6, and following the diagonal line for 1 μg of analyte, we find that the analysis of an analyte present at a concentration of 10% w/w requires a sample of only 10 μg! Extending such an analysis to a major analyte, therefore, requires the ability to obtain and work with very small samples or the ability to dilute the original sample accurately. Using this example, analyzing a sample for an analyte whose concentration is 10% w/w requires a 10,000-fold dilution. Not surprisingly, concentration methods are most commonly used for minor, trace, and ultratrace analytes, in macro and meso samples.

3D.7 Equipment, Time, and Cost

Finally, analytical methods can be compared in terms of their need for equipment, the time required to complete an analysis, and the cost per sample. Methods relying on instrumentation are equipment-intensive and may require significant operator training. For example, the graphite furnace atomic absorption spectroscopic method for determining lead levels in water requires a significant capital investment in the instrument and an experienced operator to obtain reliable results. Other methods, such as titrimetry, require only simple equipment and reagents and can be learned quickly.

The time needed to complete an analysis for a single sample is often fairly similar from method to method. This is somewhat misleading, however, because much of this time is spent preparing the solutions and equipment needed for the analysis. Once the solutions and equipment are in place, the number of samples that can be analyzed per hour differs substantially from method to method. This is a significant factor in selecting a method for laboratories that handle a high volume of samples.

The cost of an analysis is determined by many factors, including the cost of necessary equipment and reagents, the cost of hiring analysts, and the number of samples that can be processed per hour. In general, methods relying on instruments cost more per sample than other methods.

3D.8 Making the Final Choice

Unfortunately, the design criteria discussed earlier are not mutually independent.[8] Working with smaller amounts of analyte or sample, or improving selectivity, often comes at the expense of precision. Attempts to minimize cost and analysis time may decrease accuracy. Selecting a specific method requires a careful balance among these design criteria. Usually, the most important design criterion is accuracy, and the best method is that capable of producing the most accurate results. When the need for results is urgent, as is often the case in clinical labs, analysis time may become the critical factor.

The best method is often dictated by the sample's properties. Analyzing a sample with a complex matrix may require a method with excellent selectivity to avoid

interferences. Samples in which the analyte is present at a trace or ultratrace concentration usually must be analyzed by a concentration method. If the quantity of sample is limited, then the method must not require large amounts of sample.

Determining the concentration of lead in drinking water requires a method that can detect lead at the parts per billion concentrations. Selectivity is also important because other metal ions are present at significantly higher concentrations. Graphite furnace atomic absorption spectroscopy is a commonly used method for determining lead levels in drinking water because it meets these specifications. The same method is also used in determining lead levels in blood, where its ability to detect low concentrations of lead using a few microliters of sample are important considerations.

3E Developing the Procedure

After selecting a method, it is necessary to develop a procedure that will accomplish the goals of the analysis. In developing the procedure, attention is given to compensating for interferences, selecting and calibrating equipment, standardizing the method, acquiring a representative sample, and validating the method.

3E.1 Compensating for Interferences

The accuracy of a method depends on its selectivity for the analyte. Even the best methods, however, may not be free from interferents that contribute to the measured signal. Potential interferents may be present in the sample itself or the reagents used during the analysis. In this section we will briefly look at how to minimize these two sources of interference.

In the absence of an interferent, the total signal measured during an analysis, S_{meas}, is a sum of the signal due to the analyte, and the signal due to the reagents, S_{reag}

$$S_{meas} = S_A + S_{reag} = kn_A + S_{reag} \qquad \text{(total analysis method)} \qquad \textbf{3.9}$$

$$S_{meas} = S_A + S_{reag} = kC_A + S_{reag} \qquad \text{(concentration method)} \qquad \textbf{3.10}$$

Without an independent determination of S_{reag}, equation 3.9 or 3.10 cannot be solved for the moles or concentration of analyte. The contribution of S_{reag} is determined by measuring the signal for a reagent or **method blank** that does not contain the sample. Consider, for example, a procedure in which a 0.1-g sample is dissolved in 100 mL of solvent. After dissolving the sample, several reagents are added, and the signal is measured. The reagent blank is prepared by omitting the sample and adding the reagents to 100 mL of solvent. When the sample is a liquid, or is in solution, an equivalent volume of an inert solvent is substituted for the sample. Once S_{reag} is known, it is easy to correct S_{meas} for the reagent's contribution to the overall signal.

method blank
A sample that contains all components of the matrix except the analyte.

Compensating for an interference in the sample's matrix is more difficult. If the identity and concentration of the interferent are known, then it can be added to the reagent blank. In most analyses, however, the identity or concentration of matrix interferents is not known, and their contribution to S_{meas} is not included in S_{reag}. Instead, the signal from the interferent is included as an additional term

$$S_{meas} = k_A n_A + k_I n_I + S_{reag} \qquad \text{(total analysis method)} \qquad \textbf{3.11}$$

$$S_{meas} = k_A C_A + k_I C_I + S_{reag} \qquad \text{(concentration method)} \qquad \textbf{3.12}$$

Solving either equation 3.11 or 3.12 for the amount of analyte can be accomplished by separating the analyte and interferent before the analysis, thus eliminating the term for the interferent. Methods for effecting this separation are discussed in Chapter 7.

Alternatively, equations 3.11 or 3.12 can be solved for the amounts of both the analyte and the interferent. To do so, however, we must obtain two independent values for S_{meas}. Using a concentration method as an example, gives two equations

$$S_{meas,1} = k_{A,1}C_A + k_{I,1}C_I + S_{reag,1}$$

$$S_{meas,2} = k_{A,2}C_A + k_{I,2}C_I + S_{reag,2}$$

that can be solved simultaneously for C_A and C_I. This treatment is general. The composition of a solution with a total of n analytes and interferents can be determined by measuring n independent signals, and solving n independent simultaneous equations of the general form of equation 3.11 or 3.12.

Example

EXAMPLE 3.3

A sample was analyzed for the concentration of two analytes, A and B, under two sets of conditions. Under condition 1, the calibration sensitivities are

$$k_{A,1} = 76 \text{ ppm}^{-1} \qquad k_{B,1} = 186 \text{ ppm}^{-1}$$

and for condition 2

$$k_{A,2} = 33 \text{ ppm}^{-1} \qquad k_{B,2} = 243 \text{ ppm}^{-1}$$

The signals under the two sets of conditions are

$$S_{meas,1} = 33.4 \qquad S_{meas,2} = 29.7$$

Determine the concentration of A and B. You may assume that S_{reag} is zero under both conditions.

SOLUTION

Using equation 3.12, we write the following simultaneous equations

$$33.4 = (76 \text{ ppm}^{-1})C_A + (186 \text{ ppm}^{-1})C_B$$

$$29.7 = (33 \text{ ppm}^{-1})C_A + (243 \text{ ppm}^{-1})C_B$$

Multiplying the first equation by the ratio 33/76 gives the two equations as

$$14.5 = (33 \text{ ppm}^{-1})C_A + (80.8 \text{ ppm}^{-1})C_B$$

$$29.7 = (33 \text{ ppm}^{-1})C_A + (243 \text{ ppm}^{-1})C_B$$

Subtracting the first equation from the second gives

$$15.2 = (162.2 \text{ ppm}^{-1})C_B$$

Solving for C_B gives the concentration of B as 0.094 ppm. Substituting this concentration back into either of the two original equations gives the concentration of A, C_A, as 0.21 ppm.

3E.2 Calibration and Standardization

Analytical chemists make a distinction between calibration and standardization.[9] **Calibration** ensures that the equipment or instrument used to measure the signal is operating correctly by using a standard known to produce an exact signal. Balances, for example, are calibrated using a standard weight whose mass can be traced to the internationally accepted platinum–iridium prototype kilogram.

calibration
The process of ensuring that the signal measured by a piece of equipment or an instrument is correct.

Standardization is the process of experimentally determining the relationship between the signal and the amount of analyte (the value of k in equations 3.1 and 3.2). For a total analysis method, standardization is usually defined by the stoichiometry of the chemical reactions responsible for the signal. For a concentration method, however, the relationship between the signal and the analyte's concentration is a theoretical function that cannot be calculated without experimental measurements. To standardize a method, the value of k is determined by measuring the signal for one or more standards, each containing a known concentration of analyte. When several standards with different concentrations of analyte are used, the result is best viewed visually by plotting S_{meas} versus the concentration of analyte in the standards. Such a plot is known as a **calibration curve.** A more detailed discussion of calibration and standardization is found in Chapter 5.

standardization
The process of establishing the relationship between the amount of analtye and a method's signal.

calibration curve
The result of a standardization showing graphically how a method's signal changes with respect to the amount of analyte.

3E.3 Sampling

Selecting an appropriate method helps ensure that an analysis is accurate. It does not guarantee, however, that the result of the analysis will be sufficient to solve the problem under investigation or that a proposed answer will be correct. These latter concerns are addressed by carefully collecting the samples to be analyzed.

A proper sampling strategy ensures that samples are representative of the material from which they are taken. Biased or nonrepresentative sampling and contamination of samples during or after their collection are two sources of sampling error that can lead to significant errors. It is important to realize that sampling errors are completely independent of analysis errors. As a result, sampling errors cannot be corrected by evaluating a reagent blank. A more detailed discussion of sampling is found in Chapter 7.

3E.4 Validation

Before a procedure can provide useful analytical information, it is necessary to demonstrate that it is capable of providing acceptable results. **Validation** is an evaluation of whether the precision and accuracy obtained by following the procedure are appropriate for the problem. In addition, validation ensures that the written procedure has sufficient detail so that different analysts or laboratories following the same procedure obtain comparable results. Ideally, validation uses a standard sample whose composition closely matches the samples for which the procedure was developed. The comparison of replicate analyses can be used to evaluate the procedure's precision and accuracy. Intralaboratory and interlaboratory differences in the procedure also can be evaluated. In the absence of appropriate standards, accuracy can be evaluated by comparing results obtained with a new method to those obtained using a method of known accuracy. Chapter 14 provides a more detailed discussion of validation techniques.

validation
The process of verifying that a procedure yields acceptable results.

3F Protocols

quality assurance and quality control Those steps taken to ensure that the work conducted in an analytical lab is capable of producing acceptable results; also known as QA/QC.

Earlier we noted that a protocol is a set of stringent written guidelines, specifying an exact procedure that must be followed if results are to be accepted by the agency specifying the protocol. Besides all the considerations taken into account when designing the procedure, a protocol also contains very explicit instructions regarding internal and external **quality assurance and quality control** (QA/QC) procedures.[10] Internal QA/QC includes steps taken to ensure that the analytical work in a given laboratory is both accurate and precise. External QA/QC usually involves a process in which the laboratory is certified by an external agency.

As an example, we will briefly outline some of the requirements in the Environmental Protection Agency's Contract Laboratory Program (CLP) protocol for the analysis of trace metals in aqueous samples by graphite furnace atomic absorption spectrophotometry. The CLP protocol (Figure 3.7) calls for daily standardization with a reagent blank and three standards, one of which is at the laboratory's contract required detection limit. The resulting calibration curve is then verified by analyzing initial calibration verification (ICV) and initial calibration blank (ICB) samples. The reported concentration of the ICV sample must fall within ±10% of the expected concentration. If the concentration falls outside this limit, the analysis must be stopped and the problem identified and corrected before continuing.

After a successful analysis of the ICV and ICB samples, standardization is reverified by analyzing a continuing calibration verification (CCV) sample and a continuing calibration blank (CCB). Results for the CCV also must be within ±10% of the expected concentration. Again, if the concentration of the CCV falls outside the established limits, the analysis must be stopped, the problem identified and corrected, and the system standardized as described earlier. The CCV and the CCB are analyzed before the first and after the last sample, and after every set of ten samples. Whenever the CCV or the CCB is unacceptable, the results for the most recent set of ten samples are discarded, the system is standardized, and the samples are reanalyzed. By following this protocol, every result is bound by successful checks on the standardization. Although not shown in Figure 3.7, the CLP also contains detailed instructions regarding the analysis of duplicate or split samples and the use of spike testing for accuracy.

3G The Importance of Analytical Methodology

The importance of analytical methodology is evident when examining the results of environmental monitoring programs. The purpose of a monitoring program is to determine the present status of an environmental system and to assess long-term trends in the quality of the system. These are broad and poorly defined goals. In many cases, such studies are initiated with little thought to the questions the data will be used to answer. This is not surprising since it can be hard to formulate questions in the absence of initial information about the system. Without careful planning, however, a poor experimental design may result in data that has little value.

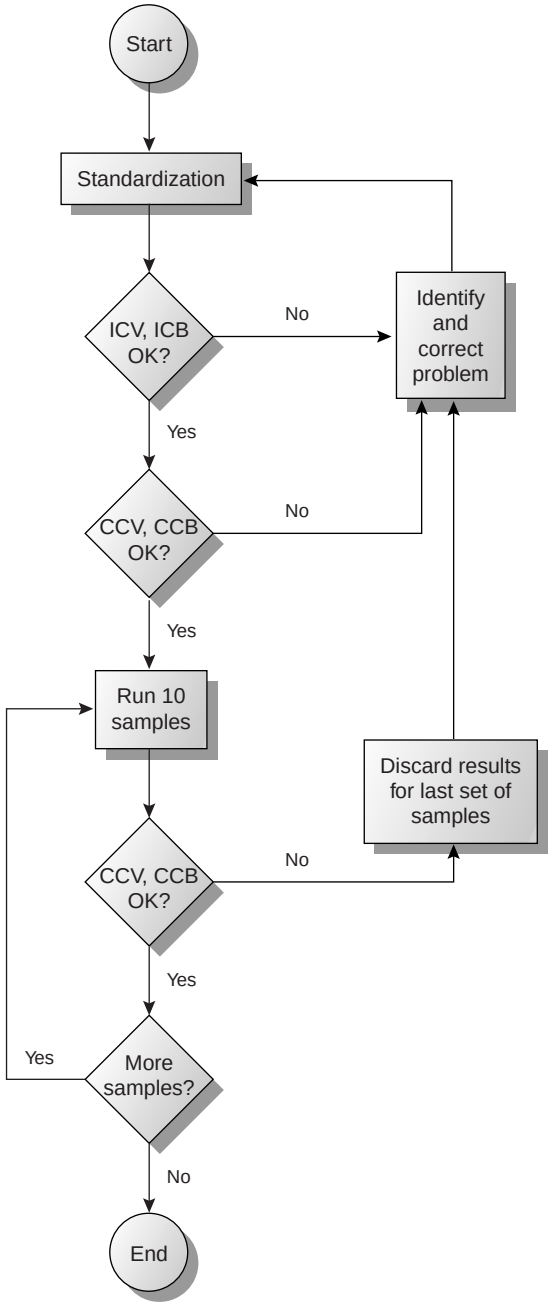


Figure 3.7
Schematic diagram of a portion of the Contract Laboratory Program protocol for the analysis of trace metals by graphite furnace atomic spectrophotometry, as specified by the Environmental Protection Agency. (*Abbreviations:* ICV = initial calibration verification; ICB = initial calibration blank, CCV = continuing calibration verification, CCB = continuing calibration blank)

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

At the beginning of the Chesapeake Bay monitoring program, little attention was given to the proper choice of analytical methods, in large part because the intended uses of the monitoring data were not specified. The analytical methods initially chosen were those standard methods already approved by the EPA. In many cases these methods proved to be of little value for this monitoring project. Most of the EPA-approved methods were designed to detect pollutants at their legally mandated maximum allowed concentrations. The concentrations of these contaminants in natural waters, however, are often well below the detection limit of the EPA methods. For example, the EPA-approved standard method for phosphate had a detection limit of 7.5 ppb. Since actual phosphate concentrations in Chesapeake Bay usually were below the EPA detection limit, the EPA method provided no useful information. On the other hand, a nonapproved variant of the EPA method commonly used in chemical oceanography had a detection limit of 0.06 ppb. In other cases, such as the elemental analysis for particulate forms of carbon, nitrogen, and phosphorus, EPA-approved procedures provided poorer reproducibility than nonapproved methods.

3H KEY TERMS

accuracy (*p. 38*)
analysis (*p. 36*)
analytes (*p. 36*)
calibration (*p. 47*)
calibration curve (*p. 47*)
concentration techniques (*p. 38*)
detection limit (*p. 39*)
determination (*p. 36*)
matrix (*p. 36*)
measurement (*p. 36*)
method (*p. 36*)
method blank (*p. 45*)
precision (*p. 39*)
procedure (*p. 36*)
protocol (*p. 37*)
quality assurance and quality control (*p. 48*)
robust (*p. 42*)
rugged (*p. 42*)
selectivity (*p. 40*)
selectivity coefficient (*p. 40*)
sensitivity (*p. 39*)
signal (*p. 37*)
standardization (*p. 47*)
technique (*p. 36*)
total analysis techniques (*p. 38*)
validation (*p. 47*)

3I SUMMARY

Every discipline has its own terminology. Your success in studying analytical chemistry will improve if you master the language used by analytical chemists. Be sure that you understand the difference between an analyte and its matrix, a technique and a method, a procedure and a protocol, and a total analysis technique and a concentration technique.

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

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

3J PROBLEMS

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

> After dissolving the sample in a beaker, remove any solid impurities by passing the solution containing the analyte through filter paper, collecting the solution in a clean Erlenmeyer flask. Rinse the beaker with several small portions of solvent, passing these rinsings through the filter paper, and collecting them in the same Erlenmeyer flask. Finally, rinse the filter paper with several portions of solvent, collecting the rinsings in the same Erlenmeyer flask.

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

> After dissolving the sample in a beaker, remove any solid impurities by filtering a portion of the solution containing the analyte. Collect and discard the first several milliliters of solution before collecting a sample of approximately 5 mL for further analysis.

Explain why these two procedures are different.

2. A certain concentration method works best when the analyte's concentration is approximately 10 ppb.
   a. If the sampling volume for the method is 0.5 mL, about what mass of analyte is being measured?
   b. If the analyte is present at 10% w/v, how would you prepare the sample for analysis?
   c. Repeat for the case in which the analyte is present at 10% w/w.
   d. Based on your results, comment on the suitability of this method for the analysis of a major analyte.

3. An analyst needs to evaluate the potential effect of an interferent, I, on the quantitative analysis for an analyte, A. She begins by measuring the signal for a sample in which the interferent is absent and the analyte is present with a concentration of 15 ppm, obtaining an average signal of 23.3 (arbitrary units). When analyzing a sample in which the analyte is absent and the interferent is present with a concentration of 25 ppm, she obtains an average signal of 13.7.
   a. What is the analyte's sensitivity?
   b. What is the interferent's sensitivity?
   c. What is the value of the selectivity coefficient?
   d. Is the method more selective for the analyte or the interferent?
   e. What is the maximum concentration of interferent relative to that of the analyte (i.e., [interferent]/[analyte]), if the error in the analysis is to be less than 1%?

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

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

6. The quantitative analysis for reduced glutathione in blood is complicated by the presence of many potential interferents. In one study, when analyzing a solution of 10-ppb glutathione and 1.5-ppb ascorbic acid, the signal was 5.43 times greater than that obtained for the analysis of 10-ppb glutathione.[12] What is the selectivity coefficient for this analysis? The same study found that when analyzing a solution of 350-ppb methionine and 10-ppb glutathione the signal was 0.906 times less than that obtained for the analysis of 10 ppb-glutathione. What is the selectivity coefficient for this analysis? In what way do these interferents behave differently?

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

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

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

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

When the concentration of glycolic acid is 1.0×10^{-4} M and the concentration of ascorbic acid is 1.0×10^{-5} M, the ratio of the two signals was found to be

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

   a. Using the ratio of the two signals, determine the value of the selectivity ratio

$$K_{GA,AA} = \frac{k_{AA}}{k_{GA}}$$

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

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

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

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

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

d. What is the largest concentration of ascorbic acid that may be present if a concentration of 1.12×10^{-6} M hypoxanthine is to be determined within ±1%?

9. A sample was analyzed for the concentration of two analytes, C and D, under two sets of conditions. Under condition 1 the calibration sensitivities are

$$k_{C,1} = 23 \text{ ppm}^{-1} \qquad k_{D,1} = 415 \text{ ppm}^{-1}$$

and for condition 2

$$k_{C,2} = 115 \text{ ppm}^{-1} \qquad k_{D,2} = 45 \text{ ppm}^{-1}$$

The signals under the two sets of conditions are

$$S_{meas,1} = 78.6 \qquad S_{meas,2} = 47.9$$

Determine the concentration of C and D. You may assume that S_{reag} is zero under both conditions.

10. Examine a procedure from *Standard Methods for the Analysis of Waters and Wastewaters* (or another manual of standard analytical methods), and identify the steps taken to compensate for interferences, to calibrate equipment and instruments, to standardize the method, and to acquire a representative sample.

3K SUGGESTED READINGS

The following papers provide alternative schemes for classifying analytical methods

Booksh, K. S.; Kowalski, B. R. "Theory of Analytical Chemistry," *Anal. Chem.* **1994,** *66,* 782A–791A.

Phillips, J. B. "Classification of Analytical Methods," *Anal. Chem.* **1981,** *53,* 1463A–1470A.

Valcárcel, M.; Luque de Castro, M. D. "A Hierarchical Approach to Analytical Chemistry," *Trends Anal. Chem.,* **1995,** *14,* 242–250.

Further details on evaluating analytical methods may be found in

Wilson, A. L. "The Performance-Characteristics of Analytical Methods," Part I-*Talanta,* **1970,** *17,* 21–29; Part II-*Talanta,* **1970,** *17,* 31–44; Part III-*Talanta,* **1973,** *20,* 725–732; Part IV-*Talanta,* **1974,** *21,* 1109–1121.

Several texts provide numerous examples of analytical procedures for specific analytes in well-defined matrices.

Basset, J.; Denney, R. C.; Jeffery, G. H.; et al. *Vogel's Textbook of Quantitative Inorganic Analysis,* 4th ed. Longman: London, 1981.

Csuros, M. *Environmental Sampling and Analysis for Technicians,* Lewis: Boca Raton, 1994.

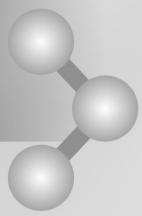
Keith, L. H., ed. *Compilation of EPA's Sampling and Analysis Methods,* Lewis: Boca Raton, 1996.

Rump, H. H.; Krist, H. *Laboratory Methods for the Examination of Water, Wastewater and Soil.* VCH Publishers: New York, 1988.

Standard Methods for the Analysis of Waters and Wastewaters, 19th ed. American Public Health Association: Washington, DC, 1995.

3L REFERENCES

1. Taylor, J. K. *Anal. Chem.* **1983,** *55,* 600A–608A.
2. Fitch, A.; Wang, Y.; Mellican, S.; et al. *Anal. Chem.* **1996,** *68,* 727A–731A.
3. Basset, J.; Denney, R. C.; Jeffery, G. H.; et al. *Vogel's Textbook of Quantitative Inorganic Analysis,* 4th ed. Longman: London, 1981, p. 8.
4. Ingle, J. D.; Crouch, S. R. *Spectrochemical Analysis.* Prentice-Hall: Englewood, NJ, 1988, pp. 171–172.
5. Barnett, N. W.; Bowser, T. A.; Gerardi, R. D.; et al. *Anal. Chim. Acta* **1996,** *318,* 309–317.
6. Rogers, L. B. *J. Chem. Ed.* **1986,** *63,* 3–6.
7. (a) Sandell, E. B.; Elving, P. J. In Kolthoff, I. M.; Elving, P. J., eds. *Treatise on Analytical Chemistry,* Interscience: New York; Part 1, Vol. 1, Chapter 1, pp. 3–6; (b) Potts, L. W. *Quantitative Analysis—Theory and Practice.* Harper and Row: New York, 1987, p. 12.
8. Valcárcel, M.; Ríos, A. *Anal. Chem.* **1993,** *65,* 781A–787A.
9. Valcárcel, M.; Ríos, A. *Analyst,* **1995,** *120,* 2291–2297.
10. (a) Amore, F. *Anal. Chem.* **1979,** *51,* 1105A–1110A; (b) Taylor, J. K. *Anal. Chem.* **1981,** *53,* 1588A–1593A.
11. D'Elia, C. F.; Sanders, J. G.; Capone, D. G. *Environ. Sci. Technol.* **1989,** *23,* 768–774.
12. Jiménez-Prieto, R.; Velasco, A.; Silva, M.; et al. *Anal. Chim. Acta* **1992,** *269,* 273–279.
13. Oungpipat, W.; Alexander, P. W. *Anal. Chim. Acta* **1994,** *295,* 36–46.
14. Ibrahim, M. S.; Ahmad, M. E.; Temerk, Y. M.; et al. *Anal. Chim. Acta* **1996,** *328,* 47–52.



Chapter 4

Evaluating Analytical Data

A problem dictates the requirements we place on our measurements and results. Regulatory agencies, for example, place stringent requirements on the reliability of measurements and results reported to them. This is the rationale for creating a protocol for regulatory problems. Screening the products of an organic synthesis, on the other hand, places fewer demands on the reliability of measurements, allowing chemists to customize their procedures.

When designing and evaluating an analytical method, we usually make three separate considerations of experimental error.[1] First, before beginning an analysis, errors associated with each measurement are evaluated to ensure that their cumulative effect will not limit the utility of the analysis. Errors known or believed to affect the result can then be minimized. Second, during the analysis the measurement process is monitored, ensuring that it remains under control. Finally, at the end of the analysis the quality of the measurements and the result are evaluated and compared with the original design criteria. This chapter is an introduction to the sources and evaluation of errors in analytical measurements, the effect of measurement error on the result of an analysis, and the statistical analysis of data.

4A Characterizing Measurements and Results

Let's begin by choosing a simple quantitative problem requiring a single measurement. The question to be answered is—What is the mass of a penny? If you think about how we might answer this question experimentally, you will realize that this problem is too broad. Are we interested in the mass of United State pennies or Canadian pennies, or is the difference in country of importance? Since the composition of a penny probably differs from country to country, let's limit our problem to pennies minted in the United States. There are other considerations. Pennies are minted at several locations in the United States (this is the meaning of the letter, or absence of a letter, below the date stamped on the lower right corner of the face of the coin). Since there is no reason to expect a difference between where the penny was minted, we will choose to ignore this consideration. Is there a reason to expect a difference between a newly minted penny not yet in circulation, and a penny that has been in circulation? The answer to this is not obvious. Let's simplify the problem by narrowing the question to—What is the mass of an average United States penny in circulation? This is a problem that we might expect to be able to answer experimentally.

A good way to begin the analysis is to acquire some preliminary data. Table 4.1 shows experimentally measured masses for seven pennies from my change jar at home. Looking at these data, it is immediately apparent that our question has no simple answer. That is, we cannot 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 these data by providing a measure of the spread of the individual measurements around a central value.

4A.1 Measures of Central Tendency

One way to characterize the data in Table 4.1 is to assume that the masses of individual pennies are scattered around a central value that provides the best estimate of a penny's true mass. Two common ways to report this estimate of central tendency are the mean and the median.

mean
The average value of a set of data ($\overline{X}$).

Mean The **mean,** $\overline{X}$, is the numerical average obtained by dividing the sum of the individual measurements by the number of measurements

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

where X_i is the i^{th} measurement, and n is the number of independent measurements.

Table 4.1 Masses of Seven United States Pennies in Circulation

Penny	Mass (g)
1	3.080
2	3.094
3	3.107
4	3.056
5	3.112
6	3.174
7	3.198

Example

EXAMPLE 4.1

What is the mean for the data in Table 4.1?

SOLUTION

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

$$3.080 + 3.094 + 3.107 + 3.056 + 3.112 + 3.174 + 3.198 = 21.821$$

and divide by the number of measurements

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

The mean is the most common estimator of central tendency. It is not considered a robust estimator, however, because extreme measurements, those much larger or smaller than the remainder of the data, strongly influence the mean's value.[2] For example, mistakenly recording the mass of the fourth penny as 31.07 g instead of 3.107 g, changes the mean from 3.117 g to 7.112 g!

Median The **median,** X_{med}, is the middle value when data are ordered from the smallest to the largest value. When the data include an odd number of measurements, the median is the middle value. For an even number of measurements, the median is the average of the $n/2$ and the $(n/2) + 1$ measurements, where n is the number of measurements.

median
That value for a set of ordered data, for which half of the data is larger in value and half is smaller in value ($\overline{X}_{med}$).

Example

EXAMPLE 4.2

What is the median for the data in Table 4.1?

SOLUTION

To determine the median, we order the data from the smallest to the largest value

3.056 3.080 3.094 3.107 3.112 3.174 3.198

Since there is a total of seven measurements, the median is the fourth value in the ordered data set; thus, the median is 3.107.

As shown by Examples 4.1 and 4.2, the mean and median provide similar estimates of central tendency when all data are similar in magnitude. The median, however, provides a more robust estimate of central tendency since it is less sensitive to measurements with extreme values. For example, introducing the transcription error discussed earlier for the mean only 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 true mass, then the spread of the individual measurements must provide an estimate of the variability in the masses of individual pennies. Although spread is often defined 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 will not change the magnitude of the spread. Three common measures of spread are range, standard deviation, and variance.

range
The numerical difference between the largest and smallest values in a data set (w).

Range The **range,** w, is the difference between the largest and smallest values in the data set.

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

The range provides information about the total variability in the data set, but does not provide any information about the distribution of individual measurements. The range for the data in Table 4.1 is the difference between 3.198 g and 3.056 g; thus

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

standard deviation
A statistical measure of the "average" deviation of data from the data's mean value (s).

Standard Deviation The absolute **standard deviation,** s, describes the spread of individual measurements about the mean and is given as

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

where X_i is one of n individual measurements, and $\overline{X}$ is the mean. Frequently, the relative standard deviation, s_r, is reported.

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

The percent relative standard deviation is obtained by multiplying s_r by 100%.

Example

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?

SOLUTION

To calculate the standard deviation, we obtain the difference between the mean value (3.117; see Example 4.1) and each measurement, square the resulting differences, and add them to determine the sum of the squares (the numerator of equation 4.1)

$$\begin{aligned}
(3.080 - 3.117)^2 &= (-0.037)^2 = 0.00137\\
(3.094 - 3.117)^2 &= (-0.023)^2 = 0.00053\\
(3.107 - 3.117)^2 &= (-0.010)^2 = 0.00010\\
(3.056 - 3.117)^2 &= (-0.061)^2 = 0.00372\\
(3.112 - 3.117)^2 &= (-0.005)^2 = 0.00003\\
(3.174 - 3.117)^2 &= (+0.057)^2 = 0.00325\\
(3.198 - 3.117)^2 &= (+0.081)^2 = \underline{0.00656}\\
&\qquad\qquad\quad\ 0.01556
\end{aligned}$$

The standard deviation is calculated by dividing the sum of the squares by $n - 1$, where n is the number of measurements, and taking the square root.

$$s = \sqrt{\frac{0.01556}{7-1}} = 0.051$$

The relative standard deviation and percent relative standard deviation are

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

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

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

Variance Another common measure of spread is the square of the standard deviation, or the **variance.** The standard deviation, rather than the variance, is usually reported because the units for standard deviation are the same as that for the mean value.

variance
The square of the standard deviation (s^2).

Example

EXAMPLE 4.4

What is the variance for the data in Table 4.1?

SOLUTION

The variance is just the square of the absolute standard deviation. Using the standard deviation found in Example 4.3 gives the variance as

$$\text{Variance} = s^2 = (0.051)^2 = 0.0026$$

4B Characterizing Experimental Errors

Realizing that our data for the mass of a penny can be characterized by a measure of central tendency and a measure of spread suggests two questions. First, does our measure of central tendency agree with the true, or expected value? Second, why are our data scattered around the central value? Errors associated with central tendency reflect the accuracy of the analysis, but the precision of the analysis is determined by those errors associated with the spread.

4B.1 Accuracy

Accuracy is a measure of how close a measure of central tendency is to the true, or expected value, μ.† Accuracy is usually expressed as either an absolute error

$$E = \overline{X} - \mu \qquad \textbf{4.2}$$

or a percent relative error, E_r.

$$E_r = \frac{\overline{X} - \mu}{\mu} \times 100 \qquad \textbf{4.3}$$

*Many scientific calculators include two keys for calculating the standard deviation, only one of which corresponds to equation 4.3. Your calculator's manual will help you determine the appropriate key to use.

†The standard convention for representing experimental parameters is to use a Roman letter for a value calculated from experimental data, and a Greek letter for the corresponding true value. For example, the experimentally determined mean is $\overline{X}$, and its underlying true value is μ. Likewise, the standard deviation by experiment is given the symbol s, and its underlying true value is identified as σ.

Although the mean is used as the measure of central tendency in equations 4.2 and 4.3, the median could also be used.

Errors affecting the accuracy of an analysis are called determinate and are characterized by a systematic deviation from the true value; that is, all the individual measurements are either too large or too small. A positive **determinate error** results in a central value that is larger than the true value, and a negative determinate error leads to a central value that is smaller than the true value. Both positive and negative determinate errors may affect the result of an analysis, with their cumulative effect leading to a net positive or negative determinate error. It is possible, although not likely, that positive and negative determinate errors may be equal, resulting in a central value with no net determinate error.

determinate error
Any systematic error that causes a measurement or result to always be too high or too small; can be traced to an identifiable source.

Determinate errors may be divided into four categories: sampling errors, method errors, measurement errors, and personal errors.

Sampling Errors We introduce determinate **sampling errors** when our sampling strategy fails to provide a representative sample. This is especially important when sampling **heterogeneous** materials. For example, determining the environmental quality of a lake by sampling a single location near a point source of pollution, such as an outlet for industrial effluent, gives misleading results. In determining the mass of a U.S. penny, the strategy for selecting pennies must ensure that pennies from other countries are not inadvertently included in the sample. Determinate errors associated with selecting a sample can be minimized with a proper sampling strategy, a topic that is considered in more detail in Chapter 7.

sampling error
An error introduced during the process of collecting a sample for analysis.

heterogeneous
Not uniform in composition.

Method Errors Determinate **method errors** are introduced when assumptions about the relationship between the signal and the analyte are invalid. In terms of the general relationships between the measured signal and the amount of analyte

$$S_{meas} = kn_A + S_{reag} \qquad \text{(total analysis method)} \qquad \textbf{4.4}$$

$$S_{meas} = kC_A + S_{reag} \qquad \text{(concentration method)} \qquad \textbf{4.5}$$

method errors exist when the sensitivity, k, and the signal due to the reagent blank, S_{reag}, are incorrectly determined. For example, methods in which S_{meas} is the mass of a precipitate containing the analyte (gravimetric method) assume that the sensitivity is defined by a pure precipitate of known stoichiometry. When this assumption fails, a determinate error will exist. Method errors involving sensitivity are minimized by standardizing the method, whereas method errors due to interferents present in reagents are minimized by using a proper reagent blank. Both are discussed in more detail in Chapter 5. Method errors due to interferents in the sample cannot be minimized by a reagent blank. Instead, such interferents must be separated from the analyte or their concentrations determined independently.

method error
An error due to limitations in the analytical method used to analyze a sample.

Measurement Errors Analytical instruments and equipment, such as glassware and balances, are usually supplied by the manufacturer with a statement of the item's maximum **measurement error,** or **tolerance.** For example, a 25-mL volumetric flask might have a maximum error of ±0.03 mL, meaning that the actual volume contained by the flask lies within the range of 24.97–25.03 mL. Although expressed as a range, the error is determinate; thus, the flask's true volume is a fixed value within the stated range. A summary of typical measurement errors for a variety of analytical equipment is given in Tables 4.2–4.4.

measurement error
An error due to limitations in the equipment and instruments used to make measurements.

tolerance
The maximum determinate measurement error for equipment or instrument as reported by the manufacturer.

Table 4.2 Measurement Errors for Selected Glassware[a]

Glassware	Volume (mL)	Measurement Errors for Class A Glassware (±mL)	Measurement Errors for Class B Glassware (±mL)
Transfer Pipets	1	0.006	0.012
	2	0.006	0.012
	5	0.01	0.02
	10	0.02	0.04
	20	0.03	0.06
	25	0.03	0.06
	50	0.05	0.10
Volumetric Flasks	5	0.02	0.04
	10	0.02	0.04
	25	0.03	0.06
	50	0.05	0.10
	100	0.08	0.16
	250	0.12	0.24
	500	0.20	0.40
	1000	0.30	0.60
	2000	0.50	1.0
Burets	10	0.02	0.04
	25	0.03	0.06
	50	0.05	0.10

[a]Specifications for class A and class B glassware are taken from American Society for Testing and Materials E288, E542 and E694 standards.

Table 4.3 Measurement Errors for Selected Balances

Balance	Capacity (g)	Measurement Error
Precisa 160M	160	±1 mg
A & D ER 120M	120	±0.1 mg
Metler H54	160	±0.01 mg

Table 4.4 Measurement Errors for Selected Digital Pipets

Pipet Range	Volume (mL or μL)[a]	Measurement Error (±%)
10–100 μL[b]	10	1.0
	50	0.6
	100	0.6
200–1000 μL[c]	200	1.5
	1000	0.8
1–10 mL[d]	1	0.6
	5	0.4
	10	0.3

[a]Units for volume same as for pipet range.

[b]Data for Eppendorf Digital Pipet 4710.

[c]Data for Oxford Benchmate.

[d]Data for Eppendorf Maxipetter 4720 with Maxitip P.

Volumetric glassware is categorized by class. Class A glassware is manufactured to comply with tolerances specified by agencies such as the National Institute of Standards and Technology. Tolerance levels for class A glassware are small enough that such glassware normally can be used 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.

Determinate measurement errors can be minimized by calibration. A pipet can be calibrated, for example, by determining the mass of water that it delivers and using the density of water to calculate the actual volume delivered by the pipet. Although glassware and instrumentation can be calibrated, it is never safe to assume that the calibration will remain unchanged during an analysis. Many instruments, in particular, drift out of calibration over time. This complication can be minimized by frequent recalibration.

personal error
An error due to biases introduced by the analyst.

Personal Errors Finally, analytical work is always subject to a variety of **personal errors,** which can include the ability to see a change in the color of an indicator used to signal the end point of a titration; biases, such as consistently overestimating or underestimating the value on an instrument's readout scale; failing to calibrate glassware and instrumentation; and misinterpreting procedural directions. Personal errors can be minimized with proper care.

Identifying Determinate Errors Determinate errors can be difficult to detect. Without knowing the true value for an analysis, the usual situation in any analysis with meaning, there is no accepted value with which the experimental result can be compared. Nevertheless, a few strategies can be used to discover the presence of a determinate error.

constant determinate error
A determinate error whose value is the same for all samples.

Some determinate errors can be detected experimentally by analyzing several samples of different size. The magnitude of a **constant determinate error** is the same for all samples and, therefore, is more significant when analyzing smaller samples. The presence of a constant determinate error can be detected by running several analyses using different amounts of sample, and looking for a systematic change in the property being measured. For example, consider a quantitative analysis in which we separate the analyte from its matrix and determine the analyte's mass. Let's assume that the sample is 50.0% w/w analyte; thus, if we analyze a 0.100-g sample, the analyte's true mass is 0.050 g. The first two columns of Table 4.5 give the true mass of analyte for several additional samples. If the analysis has a positive constant determinate error of 0.010 g, then the experimentally determined mass for

Table 4.5 Effect of Constant Positive Determinate Error on Analysis of Sample Containing 50% Analyte (%w/w)

Mass Sample (g)	True Mass of Analyte (g)	Constant Error (g)	Mass of Analyte Determined (g)	Percent Analyte Reported (%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
1.000	0.500	0.010	0.510	51.0

any sample will always be 0.010 g, larger than its true mass (column four of Table 4.5). The analyte's reported weight percent, which is shown in the last column of Table 4.5, becomes larger when we analyze smaller samples. A graph of % w/w analyte versus amount of sample shows a distinct upward trend for small amounts of sample (Figure 4.1). A smaller concentration of analyte is obtained when analyzing smaller samples in the presence of a constant negative determinate error.

A **proportional determinate error,** in which the error's magnitude depends on the amount of sample, is more difficult to detect since the result of an 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. In terms of equations 4.4 and 4.5, the reagent blank, S_{reag}, is an example of a constant determinate error, and the sensitivity, k, may be affected by proportional errors.

proportional determinate error
A determinate error whose value depends on the amount of sample analyzed.

Potential determinate errors also can be identified by analyzing a standard sample containing a known amount of analyte in a matrix similar to that of the samples being analyzed. Standard samples 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. For example, Figure 4.2 shows an analysis sheet for a typical reference material. Alternatively, the sample can be analyzed by an independent method known to give accurate results, and the results of the two methods can be compared. Once identified, the source of a determinate error can be corrected. The best prevention against errors affecting accuracy, however, is a well-designed procedure that identifies likely sources of determinate errors, coupled with careful laboratory work.

standard reference material
A material available from the National Institute of Standards and Technology certified to contain known concentrations of analytes.

The data in Table 4.1 were obtained using a calibrated balance, certified by the manufacturer to have a tolerance of less than ±0.002 g. Suppose the Treasury Department reports that the mass of a 1998 U.S. penny is approximately 2.5 g. Since the mass of every penny in Table 4.1 exceeds the reported mass by an amount significantly greater than the balance's tolerance, we can safely conclude that the error in this analysis is not due to equipment error. The actual source of the error is revealed later in this chapter.

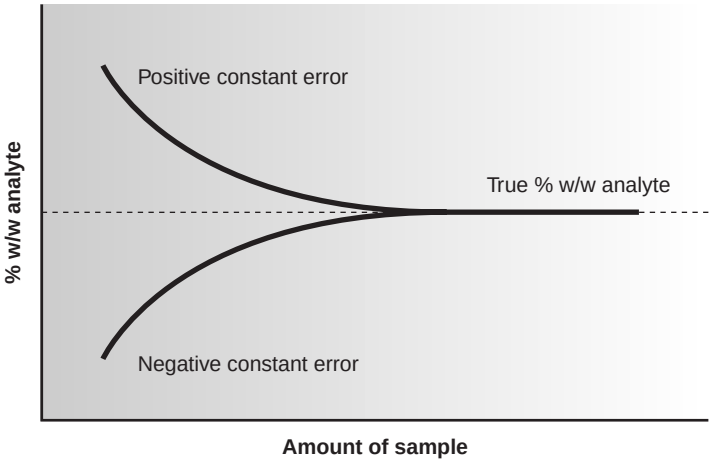


Figure 4.1
Effect of a constant determinate error on the reported concentration of analyte.

Table 4.6 Effect of Proportional Positive Determinate Error on Analysis of Sample Containing 50% Analyte (%w/w)

Mass Sample (g)	True Mass of Analyte (g)	Proportional Error (%)	Mass of Analyte Determined (g)	Percent Analyte Reported (%w/w)
0.200	0.100	1.00	0.101	50.5
0.400	0.200	1.00	0.202	50.5
0.600	0.300	1.00	0.303	50.5
0.800	0.400	1.00	0.404	50.5
1.000	0.500	1.00	0.505	50.5

Simulated Rainwater (liquid form)

This SRM was developed to aid in the analysis of acidic rainwater by providing a stable, homogeneous material at two levels of acidity.

SRM	Type	Unit of issue	
2694a	Simulated rainwater	Set of 4: 2 of 50 mL at each of 2 levels	
	Constituent element parameter	**2694a-I**	**2694a-II**
	pH, 25°C	4.30	3.60
	Electrolytic Conductivity (S/cm, 25°C)	25.4	129.3
	Acidity, meq/L	0.0544	0.283
	Fluoride, mg/L	0.057	0.108
	Chloride, mg/L	(0.23)*	(0.94)*
	Nitrate, mg/L	(0.53)*	7.19
	Sulfate, mg/L	2.69	10.6
	Sodium, mg/L	0.208	0.423
	Potassium, mg/L	0.056	0.108
	Ammonium, mg/L	(0.12)*	(1.06)*
	Calcium, mg/L	0.0126	0.0364
	Magnesium, mg/L	0.0242	0.0484

* Values in parentheses are not certified and are given for information only.

Figure 4.2
Analysis sheet for Simulated Rainwater (SRM 2694a). Adapted from NIST Special Publication 260: *Standard Reference Materials Catalog 1995–96,* p. 64; U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology.

4B.2 Precision

Precision is a measure of the spread of data about a central value and may be expressed as the range, the standard deviation, or the variance. Precision is commonly divided into two categories: repeatability and reproducibility. **Repeatability** is the precision obtained when all measurements are made by the same analyst during a single period of laboratory work, using the same solutions and equipment. **Reproducibility,** on the other hand, is the precision obtained under any other set of conditions, including that between analysts, or between laboratory sessions for a single analyst. Since reproducibility includes additional sources of variability, the reproducibility of an analysis can be no better than its repeatability.

repeatability
The precision for an analysis in which the only source of variability is the analysis of replicate samples.

reproducibility
The precision when comparing results for several samples, for several analysts or several methods.

indeterminate error
Any random error that causes some measurements or results to be too high while others are too low.

Errors affecting the distribution of measurements around a central value are called indeterminate and are characterized by a random variation in both magnitude and direction. **Indeterminate errors** need not affect the accuracy of an analysis. Since indeterminate errors are randomly scattered around a central value, positive and negative errors tend to cancel, 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 Indeterminate errors can be traced to several sources, including the collection of samples, the manipulation of samples during the analysis, and the making of measurements.

When collecting a sample, for instance, only a small portion of the available material is taken, increasing the likelihood that small-scale inhomogeneities in the sample will affect the repeatability of the analysis. Individual pennies, for example, are expected to show variation 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 error associated with the sampling process.

During the analysis numerous opportunities arise for random variations in the way individual samples are treated. In determining the mass of a penny, for example, each penny should be handled in the same manner. Cleaning some pennies but not cleaning others introduces an indeterminate error.

Finally, any measuring device is subject to an indeterminate error in reading its scale, with the last digit always being an estimate subject to random fluctuations, or background noise. For example, a buret with scale divisions every 0.1 mL has an inherent indeterminate error of ±0.01 – 0.03 mL when estimating the volume to the hundredth of a milliliter (Figure 4.3). Background noise in an electrical meter (Figure 4.4) can be evaluated by recording the signal without analyte and observing the fluctuations in the signal over time.

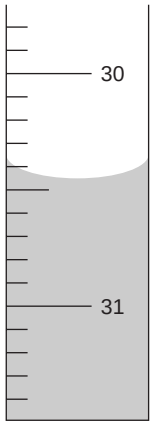


Figure 4.3
Close-up of buret, showing difficulty in estimating volume. With scale divisions every 0.1 mL it is difficult to read the actual volume to better than ±0.01 – 0.03 mL.

Evaluating Indeterminate Error Although it is impossible to eliminate indeterminate error, its effect can be minimized if the sources and relative magnitudes of the indeterminate error are known. Indeterminate errors may be estimated by an appropriate measure of spread. Typically, a standard deviation is used, although in some cases estimated values are used. The contribution from analytical instruments and equipment are easily measured or estimated. Indeterminate errors introduced by the analyst, such as inconsistencies in the treatment of individual samples, are more difficult to estimate.

To evaluate the effect of indeterminate error on the data in Table 4.1, ten replicate determinations of the mass of a single penny were made, with results shown in Table 4.7. The standard deviation for the data in Table 4.1 is 0.051, and it is 0.0024 for the data in Table 4.7. The significantly better precision when determining the mass of a single penny suggests that the precision of this analysis is not limited by the balance used to measure mass, but is due to a significant variability in the masses of individual pennies.

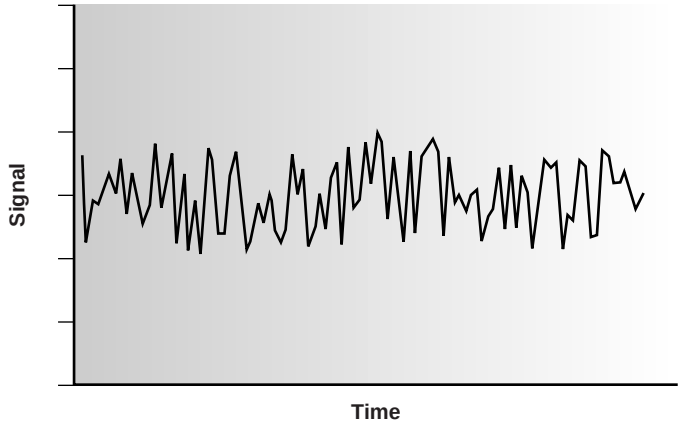


Figure 4.4
Background noise in a meter obtained by measuring signal over time in the absence of analyte.

Table 4.7 Replicate Determinations of the Mass of a Single United States Penny in Circulation

Replicate Number	Mass (g)
1	3.025
2	3.024
3	3.028
4	3.027
5	3.028
6	3.023
7	3.022
8	3.021
9	3.026
10	3.024

4B.3 Error and Uncertainty

error
A measure of bias in a result or measurement.

Analytical chemists make a distinction between error and uncertainty.[3] **Error** is the difference between a single measurement or result and its true value. In other words, error is a measure of bias. As discussed earlier, error can be divided into determinate and indeterminate sources. Although we can correct for determinate error, the indeterminate portion of the error remains. Statistical significance testing, which is discussed later in this chapter, provides a way to determine whether a bias resulting from determinate error might be present.

uncertainty
The range of possible values for a measurement.

Uncertainty expresses the range of possible values that a measurement or result might reasonably be expected to have. Note that this definition of uncertainty is not the same as that for precision. The precision of an analysis, whether reported as a range or a standard deviation, is calculated from experimental data and provides an estimation of indeterminate error affecting measurements. Uncertainty accounts for all errors, both determinate and indeterminate, that might affect our result. Although we always try to correct determinate errors, the correction itself is subject to random effects or indeterminate errors.

Table 4.8 Experimentally Determined Volumes Delivered by a 10-mL Class A Pipet

Trial	Volume Delivered (mL)	Trial	Volume Delivered (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

To illustrate the difference between precision and uncertainty, consider the use of a class A 10-mL pipet for delivering solutions. A pipet's uncertainty is the range of volumes in which its true volume is expected to lie. Suppose you purchase a 10-mL class A pipet from a laboratory supply company and use it without calibration. The pipet's tolerance value of ±0.02 mL (see Table 4.2) represents your uncertainty since your best estimate of its volume is 10.00 mL ±0.02 mL. Precision is determined experimentally by using the pipet several times, measuring the volume of solution delivered each time. Table 4.8 shows results for ten such trials that have a mean of 9.992 mL and a standard deviation of 0.006. This standard deviation represents the precision with which we expect to be able to deliver a given solution using any class A 10-mL pipet. In this case the uncertainty in using a pipet is worse than its precision. Interestingly, the data in Table 4.8 allow 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 true volume, then the uncertainty is ±0.006. As expected, calibrating the pipet allows us to lower its uncertainty.

4C Propagation of Uncertainty

Suppose that you need to add a reagent to a flask by several successive transfers using a class A 10-mL pipet. By calibrating the pipet (see Table 4.8), you know that it delivers a volume of 9.992 mL with a standard deviation of 0.006 mL. Since the pipet is calibrated, we can use the standard deviation as a measure of uncertainty. This uncertainty tells us that when we use the pipet to repetitively deliver 10 mL of solution, the volumes actually delivered are randomly scattered around the mean of 9.992 mL.

If the uncertainty in using the pipet once is 9.992 ± 0.006 mL, what is the uncertainty when the pipet is used twice? As a first guess, we might simply add the uncertainties 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 see 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 both volumes are either greater than 9.992 mL or less than 9.992 mL. At the other extreme, we might assume that the two deliveries will always be on opposite sides of the pipet's mean volume. In this case we subtract the uncertainties for the two deliveries,

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

underestimating the total uncertainty.

So what is the total uncertainty when using this pipet to deliver two successive volumes of solution? From the previous discussion we know that the total uncertainty is greater than ±0.000 mL and less than ±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 that we will not derive. A more thorough treatment can be found elsewhere.[4]

4C.1 A Few Symbols

Propagation of uncertainty allows us to estimate the uncertainty in a calculated result from the uncertainties of the measurements used to calculate the result. In the equations presented in this section the result is represented by the symbol R and the measurements by the symbols A, B, and C. The corresponding uncertainties are s_R, s_A, s_B, and s_C. The uncertainties for A, B, and C can be reported in several ways, including calculated standard deviations or estimated ranges, as long as the same form is used for all measurements.

4C.2 Uncertainty When Adding or Subtracting

When measurements are added or subtracted, the absolute uncertainty in the result is the square root of the sum of the squares of the absolute uncertainties for the individual measurements. Thus, for the equations $R = A + B + C$ or $R = A + B - C$, or any other combination of adding and subtracting A, B, and C, the absolute uncertainty in R is

$$s_R = \sqrt{s_A^2 + s_B^2 + s_C^2} \qquad \textbf{4.6}$$

Example

EXAMPLE 4.5

The class A 10-mL pipet characterized in Table 4.8 is used to deliver two successive volumes. Calculate the absolute and relative uncertainties for the total delivered volume.

SOLUTION

The total delivered volume is obtained by adding the volumes of each delivery; thus

$$V_{tot} = 9.992\ \text{mL} + 9.992\ \text{mL} = 19.984\ \text{mL}$$

Using the standard deviation as an estimate of uncertainty, the uncertainty in the total delivered volume is

$$s_R = \sqrt{(0.006)^2 + (0.006)^2} = 0.0085$$

Thus, we report the volume and its absolute uncertainty as 19.984 ± 0.008 mL. The relative uncertainty in the total delivered volume is

$$\frac{0.0085}{19.984} \times 100 = 0.043\%$$

4C.3 Uncertainty When Multiplying or Dividing

When measurements are multiplied or divided, the relative uncertainty in the result is the square root of the sum of the squares of the relative uncertainties for the individual measurements. Thus, for the equations $R = A \times B \times C$ or $R = A \times B/C$, or any other combination of multiplying and dividing A, B, and C, the relative uncertainty in R is

$$\frac{s_R}{R} = \sqrt{\left(\frac{s_A}{A}\right)^2 + \left(\frac{s_B}{B}\right)^2 + \left(\frac{s_C}{C}\right)^2} \qquad \textbf{4.7}$$

Example

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 ± 0.01 A passes through the circuit for 120 ± 1 s, the total charge is

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

Calculate the absolute and relative uncertainties for the total charge.

SOLUTION

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

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

or ±6.7%. The absolute uncertainty in the charge is

$$s_R = R \times 0.0672 = (18) \times (\pm 0.0672) = \pm 1.2$$

Thus, we report the total charge as 18 C ± 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, the propagation of uncertainty is easily calculated by treating each operation separately using equations 4.6 and 4.7 as needed.

Example

EXAMPLE 4.7

For a concentration technique the relationship between the measured signal and an analyte's concentration is given by equation 4.5

$$S_{meas} = kC_A + S_{reag}$$

Calculate the absolute and relative uncertainties for the analyte's concentration if S_{meas} is 24.37 ± 0.02, S_{reag} is 0.96 ± 0.02, and k is 0.186 ± 0.003 ppm^{-1}.

SOLUTION

Rearranging equation 4.5 and solving for C_A

$$C_A = \frac{S_{meas} - S_{reag}}{k} = \frac{24.37 - 0.96}{0.186\ ppm^{-1}} = 125.9\ ppm$$

gives the analyte's concentration as 126 ppm. To estimate the uncertainty in C_A, we first determine the uncertainty for the numerator, $S_{meas} - S_{reag}$, using equation 4.6

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

The numerator, therefore, is 23.41 ± 0.028 (note that we retain an extra significant figure since we will use this uncertainty in further calculations). To complete the calculation, we estimate the relative uncertainty in C_A using equation 4.7, giving

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

or a percent relative uncertainty of 1.6%. The absolute uncertainty in the analyte's concentration is

$$s_R = (125.9\ ppm) \times (0.0162) = \pm 2.0\ ppm$$

giving the analyte's concentration as 126 ± 2 ppm.

4C.5 Uncertainty for Other Mathematical Functions

Many other mathematical operations are commonly used in analytical chemistry, including powers, roots, and logarithms. Equations for the propagation of uncertainty for some of these functions are shown in Table 4.9.

Example

EXAMPLE 4.8

The pH of a solution is defined as

$$pH = -\log[H^+]$$

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

SOLUTION

The molar concentration of H^+ for this pH is

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

or 1.9×10^{-4} M to two significant figures. From Table 4.9 the relative uncertainty in $[H^+]$ is

$$\frac{s_R}{R} = 2.303 \times s_A = 2.303 \times 0.03 = 0.069$$

and the absolute uncertainty is

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

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

Table 4.9 Propagation of Uncertainty for Selected Functions[a]

Function	s_R
$R = kA$	$s_R = ks_A$
$R = A + B$	$s_R = \sqrt{s_A^2 + s_B^2}$
$R = A - B$	$s_R = \sqrt{s_A^2 + s_B^2}$
$R = A \times B$	$\frac{s_R}{R} = \sqrt{\left(\frac{s_A}{A}\right)^2 + \left(\frac{s_B}{B}\right)^2}$
$R = \frac{A}{B}$	$\frac{s_R}{R} = \sqrt{\left(\frac{s_A}{A}\right)^2 + \left(\frac{s_B}{B}\right)^2}$
$R = \ln(A)$	$s_R = \frac{s_A}{A}$
$R = \log(A)$	$s_R = 0.4343 \times \frac{s_A}{A}$
$R = e^A$	$\frac{s_R}{R} = s_A$
$R = 10^A$	$\frac{s_R}{R} = 2.303 s_A$
$R = A^k$	$\frac{s_R}{R} = \left[k \frac{s_A}{A}\right]$

[a]These equations assume that the measurements A and B are uncorrelated; that is, s_A is independent of s_B.

4C.6 Is Calculating Uncertainty Actually Useful?

Given the complexity of determining a result's uncertainty when several measurements are involved, it is worth examining some of the reasons why such calculations are useful. A propagation of uncertainty allows us to estimate an ex-

pected uncertainty for an analysis. Comparing the expected uncertainty to that which is actually obtained can provide useful information. For example, in determining the mass of a penny, we estimated the uncertainty in measuring mass as ±0.002 g based on the balance's tolerance. If we measure a single penny's mass several times and obtain a standard deviation of ±0.020 g, we would have reason to believe that our measurement process is out of control. We would then try to identify and correct the problem.

A propagation of uncertainty also helps in deciding how to improve the uncertainty in an analysis. In Example 4.7, for instance, we calculated the concentration of an analyte, obtaining a value of 126 ppm with an absolute uncertainty of ±2 ppm and a relative uncertainty of 1.6%. How might we improve the analysis so that the absolute uncertainty is only ±1 ppm (a relative uncertainty of 0.8%)? Looking back on the calculation, we find that the relative uncertainty is determined by the relative uncertainty in the measured signal (corrected for the reagent blank)

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

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

$$\frac{0.003}{0.186} = \pm 0.016, \text{ or } \pm 1.6\%$$

Of these two terms, the sensitivity's uncertainty dominates the total uncertainty. Measuring the signal more carefully will not improve the overall uncertainty of the analysis. On the other hand, the desired improvement in uncertainty can be achieved if the sensitivity's absolute uncertainty can be decreased to ±0.0015 ppm^{-1}.

As a final example, a propagation of uncertainty can be used to decide which of several procedures provides the smallest overall uncertainty. Preparing a solution by diluting a stock solution can be done using several different combinations of volumetric glassware. For instance, we can dilute a 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. The same dilution also can be accomplished in two steps using a 50-mL pipet and a 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, of course, depends on the uncertainty of the glassware used in the dilutions. As shown in the following example, we can use the tolerance values for volumetric glassware to determine the optimum dilution strategy.[5]

Example

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

Letting M_a and M_b represent the molarity of the final solutions from method (a) and method (b), we can write the following equations

$$M_a = 0.0010\ \text{M} = \frac{(1.0\ \text{M})(1.000\ \text{mL})}{1000.0\ \text{mL}}$$

$$M_b = 0.0010\ \text{M} = \frac{(1.0\ \text{M})(20.00\ \text{mL})(25.00\ \text{mL})}{(1000.0\ \text{mL})(500.0\ \text{mL})}$$

Using the tolerance values for pipets and volumetric flasks given in Table 4.2, the overall uncertainties in M_a and M_b are

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

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

Since the relative uncertainty for M_b is less than that for M_a, we find that the two-step dilution provides the smaller overall uncertainty.

4D The Distribution of Measurements and Results

An analysis, particularly a quantitative analysis, is usually performed on several replicate samples. How do we report the result for such an experiment when results for the replicates are scattered around a central value? To complicate matters further, the analysis of each replicate usually requires multiple measurements that, themselves, are scattered around a central value.

Consider, for example, the data in Table 4.1 for the mass of a penny. Reporting only the mean is insufficient because it fails to indicate the uncertainty in measuring a penny's mass. Including the standard deviation, or other measure of spread, provides the necessary information about the uncertainty in measuring mass. Nevertheless, the central tendency and spread together do not provide a definitive statement about a penny's true mass. If you are not convinced that this is true, ask yourself how obtaining the mass of an additional penny will change the mean and standard deviation.

How we report the result of an experiment is further complicated by the need to compare the results of different experiments. For example, Table 4.10 shows results for a second, independent experiment to determine the mass of a U.S. penny in circulation. Although the results shown in Tables 4.1 and 4.10 are similar, they are not identical; thus, we are justified in asking whether the results are in agreement. Unfortunately, a definitive comparison between these two sets of data is not possible based solely on their respective means and standard deviations.

Developing a meaningful method for reporting an experiment's result requires the ability to predict the true central value and true spread of the population under investigation from a limited sampling of that population. In this section we will take a quantitative look at how individual measurements and results are distributed around a central value.

Table 4.10 Results for a Second Determination of the Mass of a United States Penny in Circulation

Penny	Mass (g)
1	3.052
2	3.141
3	3.083
4	3.083
5	3.048
$\bar{X}$	3.081
s	0.037

4D.1 Populations and Samples

In the previous section we introduced the terms "population" and "sample" in the context of reporting the result of an experiment. Before continuing, we need to understand the difference between a population and a sample. A **population** is the set of all objects in the system being investigated. These objects, which also are members of the population, possess qualitative or quantitative characteristics, or values, that can be measured. If we analyze every member of a population, we can determine the population's true central value, μ, and spread, σ.

population
All members of a system.

The probability of occurrence for a particular value, $P(V)$, is given as

$$P(V) = \frac{M}{N}$$

where V is the value of interest, M is the value's frequency of occurrence in the population, and N is the size of the population. In determining the mass of a circulating United States penny, for instance, the members of the population are all United States pennies currently in circulation, while the values are the possible masses that a penny may have.

In most circumstances, populations are so large that it is not feasible to analyze every member of the population. This is certainly true for the population of circulating U.S. pennies. Instead, we select and analyze a limited subset, or **sample,** of the population. The data in Tables 4.1 and 4.10, for example, give results for two samples drawn at random from the larger population of all U.S. pennies currently in circulation.

sample
Those members of a population that we actually collect and analyze.

4D.2 Probability Distributions for Populations

To predict the properties of a population on the basis of a sample, it is necessary to know something about the population's expected distribution around its central value. The distribution of a population can be represented by plotting the frequency of occurrence of individual values as a function of the values themselves. Such plots are called **probability distributions.** Unfortunately, we are rarely able to calculate the exact probability distribution for a chemical system. In fact, the probability distribution can take any shape, depending on the nature of the chemical system being investigated. Fortunately many chemical systems display one of several common probability distributions. Two of these distributions, the binomial distribution and the normal distribution, are discussed next.

probability distribution
Plot showing frequency of occurrence for members of a population.

binomial distribution
Probability distribution showing chance of obtaining one of two specific outcomes in a fixed number of trials.

Binomial Distribution The **binomial distribution** describes a population in which the values are the number of times a particular outcome occurs during a fixed number of trials. Mathematically, the binomial distribution is given as

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

where $P(X,N)$ is the probability that a given outcome will occur X times during N trials, and p is the probability that the outcome will occur in a single trial.* If you flip a coin five times, $P(2,5)$ gives the probability that two of the five trials will turn up "heads."

A binomial distribution has well-defined measures of central tendency and spread. The true mean value, for example, is given as

$$\mu = Np$$

and the true spread is given by the variance

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

or the standard deviation

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

homogeneous
Uniform in composition.

The binomial distribution describes a population whose members have only certain, discrete values. A good example of a population obeying the binomial distribution is the sampling of **homogeneous** materials. As shown in Example 4.10, the binomial distribution can be used to calculate the probability of finding a particular isotope in a molecule.

Example

EXAMPLE 4.10

Carbon has two common isotopes, ^{12}C and ^{13}C, with relative isotopic abundances of, respectively, 98.89% and 1.11%. (a) What are the mean and standard deviation for the number of ^{13}C atoms in a molecule of cholesterol? (b) What is the probability of finding a molecule of cholesterol ($C_{27}H_{44}O$) containing no atoms of ^{13}C?

SOLUTION

The probability of finding an atom of ^{13}C in cholesterol follows a binomial distribution, where X is the sought for frequency of occurrence of ^{13}C atoms, N is the number of C atoms in a molecule of cholesterol, and p is the probability of finding an atom of ^{13}C.

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

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

with a standard deviation of

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

(b) Since the mean is less than one atom of ^{13}C per molecule, most molecules of cholesterol will not have any ^{13}C. To calculate

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

the probability, we substitute appropriate values into the binomial equation

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

There is therefore a 74.0% probability that a molecule of cholesterol will not have an atom of ^{13}C.

A portion of the binomial distribution for atoms of ^{13}C in cholesterol is shown in Figure 4.5. Note in particular that there is little probability of finding more than two atoms of ^{13}C in any molecule of cholesterol.

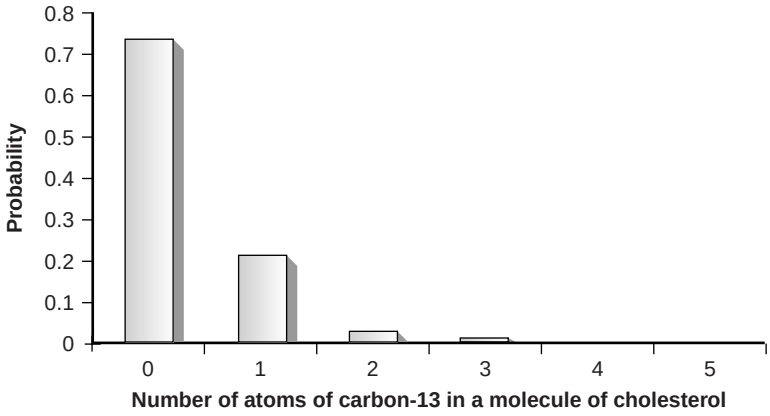


Figure 4.5
Portion of the binomial distribution for the number of naturally occurring ^{13}C atoms in a molecule of cholesterol.

Normal Distribution The binomial distribution describes a population whose members have only certain, discrete values. This is the case with the number of ^{13}C atoms in a molecule, which must be an integer number no greater then the number of carbon atoms in the molecule. A molecule, for example, cannot have 2.5 atoms of ^{13}C. Other populations are considered continuous, in that members of the population may take on any value.

The most commonly encountered continuous distribution is the Gaussian, or **normal distribution,** where the frequency of occurrence for a value, X, is given by

normal distribution
"Bell-shaped" probability distribution curve for measurements and results showing the effect of random error.

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(X-\mu)^2}{2\sigma^2}\right]$$

The shape of a normal distribution is determined by two parameters, the first of which is the population's central, or true mean value, μ, given as

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

where n is the number of members in the population. The second parameter is the population's variance, σ^2, which is calculated using the following equation*

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

*Note the difference between the equation for a population's variance, which includes the term n in the denominator, and the similar equation for the variance of a sample (the square of equation 4.3), which includes the term $n - 1$ in the denominator. The reason for this difference is discussed later in the chapter.

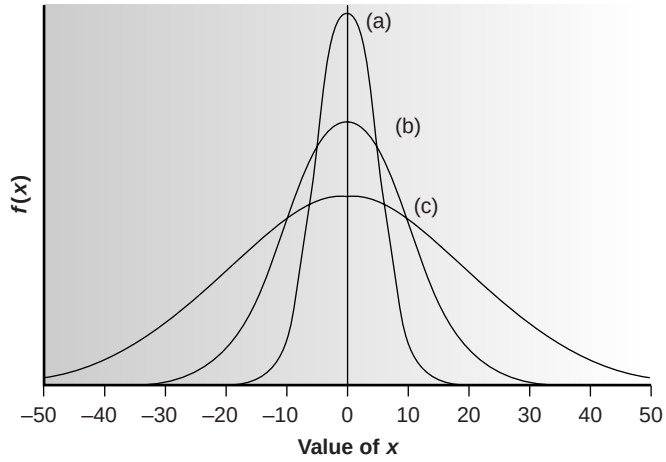


Figure 4.6
Normal distributions for (a) $\mu = 0$ and $\sigma^2 = 25$; (b) $\mu = 0$ and $\sigma^2 = 100$; and (c) $\mu = 0$ and $\sigma^2 = 400$.

Examples of normal distributions with $\mu = 0$ and $\sigma^2 = 25$, 100 or 400, are shown in Figure 4.6. Several features of these normal distributions deserve attention. First, note that each normal distribution contains a single maximum corresponding to μ and that the distribution is symmetrical about this value. Second, increasing the population's variance increases the distribution's spread while decreasing its height. Finally, because the normal distribution depends solely on μ and σ^2, the area, or probability of occurrence between any two limits defined in terms of these parameters is the same for all normal distribution curves. For example, 68.26% of the members in a normally distributed population have values within the range $\mu \pm 1\sigma$, regardless of the actual values of μ and σ. As shown in Example 4.11, probability tables (Appendix 1A) can be used to determine the probability of occurrence between any defined limits.

Example

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

The normal distribution for this example is shown in Figure 4.7, with the shaded area representing the percentage of tablets containing between 243 and 262 mg of aspirin. To determine the percentage of tablets between these limits, we first determine the percentage of tablets with less than 243 mg of aspirin, and the percentage of tablets having more than 262 mg of aspirin. This is accomplished by calculating the deviation, z, of each limit from μ, using the following equation

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

where X is the limit in question, and σ, the population standard deviation, is 5. Thus, the deviation for the lower limit is

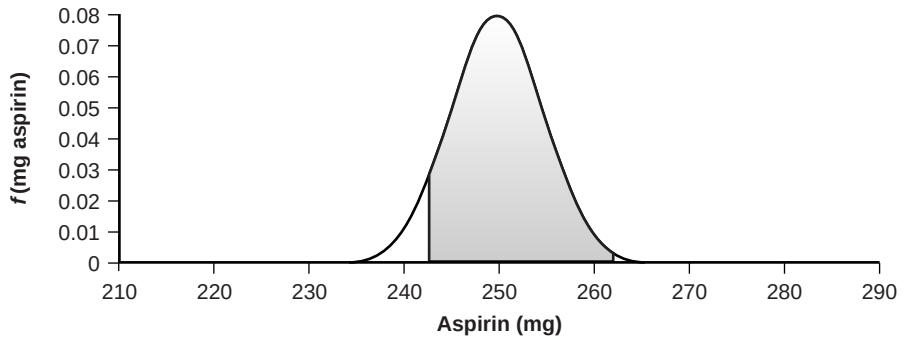


Figure 4.7
Normal distribution for population of aspirin tablets with $\mu = 250$ mg aspirin and $\sigma^2 = 25$. The shaded area shows the percentage of tablets containing between 243 and 262 mg of aspirin.

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

and the deviation for the upper limit is

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

Using the table in Appendix 1A, 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%. The percentage of tablets containing between 243 and 262 mg of aspirin is therefore

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

4D.3 Confidence Intervals for Populations

If we randomly select a single member from a population, what will be its most likely value? This is an important question, and, in one form or another, it is the fundamental problem for any analysis. One of the most important features of a population's probability distribution is that it provides a way to answer this question.

Earlier we noted that 68.26% of a normally distributed population is found within the range of $\mu \pm 1\sigma$. Stating this another way, there is a 68.26% probability that a member selected at random from a normally distributed population will have a value in the interval of $\mu \pm 1\sigma$. In general, we can write

$$X_i = \mu \pm z\sigma \qquad \textbf{4.9}$$

where the factor z accounts for the desired level of confidence. 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. Confidence intervals can be quoted for any desired probability level, several examples of which are shown in Table 4.11. For reasons that will be discussed later in the chapter, a 95% confidence interval frequently is reported.

Table 4.11 Confidence Intervals for Normal Distribution Curves Between the Limits $\mu \pm z\sigma$

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

confidence interval
Range of results around a mean value that could be explained by random error.

Example

EXAMPLE 4.12

What is the 95% confidence interval for the amount of aspirin in a single analgesic tablet drawn from a population where μ is 250 mg and σ^2 is 25?

SOLUTION

According to Table 4.11, the 95% confidence interval for a single member of a normally distributed population is

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

Thus, we expect that 95% of the tablets in the population contain between 240 and 260 mg of aspirin.

Alternatively, a confidence interval can be expressed in terms of the population's standard deviation and the value of a single member drawn from the population. Thus, equation 4.9 can be rewritten as a confidence interval for the population mean

$$\mu = X_i \pm z\sigma \qquad \textbf{4.10}$$

Example

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. A single tablet is randomly selected, analyzed, and found to contain 245 mg of aspirin. What is the 95% confidence interval for the population mean?

SOLUTION

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

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

There is, therefore, a 95% probability that the population's mean, μ, lies within the range of 231–259 mg of aspirin.

Confidence intervals also can be reported using the mean for a sample of size n, drawn from a population of known σ. The standard deviation for the mean value, $\sigma_{\bar{X}}$, 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

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

Example

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 given as

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

Thus, there is a 95% probability that the population's mean is between 239 and 251 mg of aspirin. As expected, the confidence interval based on the mean of five members of the population is smaller than that based on a single member.

4D.4 Probability Distributions for Samples

In Section 4D.2 we introduced two probability distributions commonly encountered when studying populations. The construction of confidence intervals for a normally distributed population was the subject of Section 4D.3. We have yet to address, however, how we can identify the probability distribution for a given population. In Examples 4.11–4.14 we assumed that the amount of aspirin in analgesic tablets is normally distributed. We are justified in asking how this can be determined without analyzing every member of the population. When we cannot study the whole population, or when we cannot predict the mathematical form of a population's probability distribution, we must deduce the distribution from a limited sampling of its members.

Sample Distributions and the Central Limit Theorem Let's return to the problem of determining a penny's mass to explore the relationship between a population's distribution and the distribution of samples drawn from that population. The data shown in Tables 4.1 and 4.10 are insufficient for our purpose because they are not large enough to give a useful picture of their respective probability distributions. A better picture of the probability distribution requires a larger sample, such as that shown in Table 4.12, for which $\overline{X}$ is 3.095 and s^2 is 0.0012.

The data in Table 4.12 are best displayed as a **histogram,** in which the frequency of occurrence for equal intervals of data is plotted versus the midpoint of each interval. Table 4.13 and Figure 4.8 show a frequency table and histogram for the data in Table 4.12. Note that the histogram was constructed such that the mean value for the data set is centered within its interval. In addition, a normal distribution curve using $\overline{X}$ and s^2 to estimate μ and σ^2 is superimposed on the histogram.

histogram
A plot showing the number of times an observation occurs as a function of the range of observed values.

It is noteworthy that the histogram in Figure 4.8 approximates the normal distribution curve. Although the histogram for the mass of pennies is not perfectly symmetrical, it is roughly symmetrical about the interval containing the greatest number of pennies. In addition, we know from Table 4.11 that 68.26%, 95.44%, and 99.73% of the members of a normally distributed population are within, respectively, $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$. If we assume that the mean value, 3.095 g, and the sample variance, 0.0012, are good approximations for μ and σ^2, we find that 73%,

Table 4.12 Individual Masses for a Large Sample of U.S. Pennies in Circulation[a]

Penny	Weight (g)	Penny	Weight (g)	Penny	Weight (g)	Penny	Weight (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

[a]Pennies are identified in the order in which they were sampled and weighed.

Table 4.13 Frequency Distribution for the Data in Table 4.12

Interval	Frequency
2.991–3.009	2
3.010–3.028	0
3.029–3.047	4
3.048–3.066	19
3.067–3.085	15
3.086–3.104	23
3.105–3.123	19
3.124–3.142	12
3.143–3.161	13
3.162–3.180	1
3.181–3.199	2

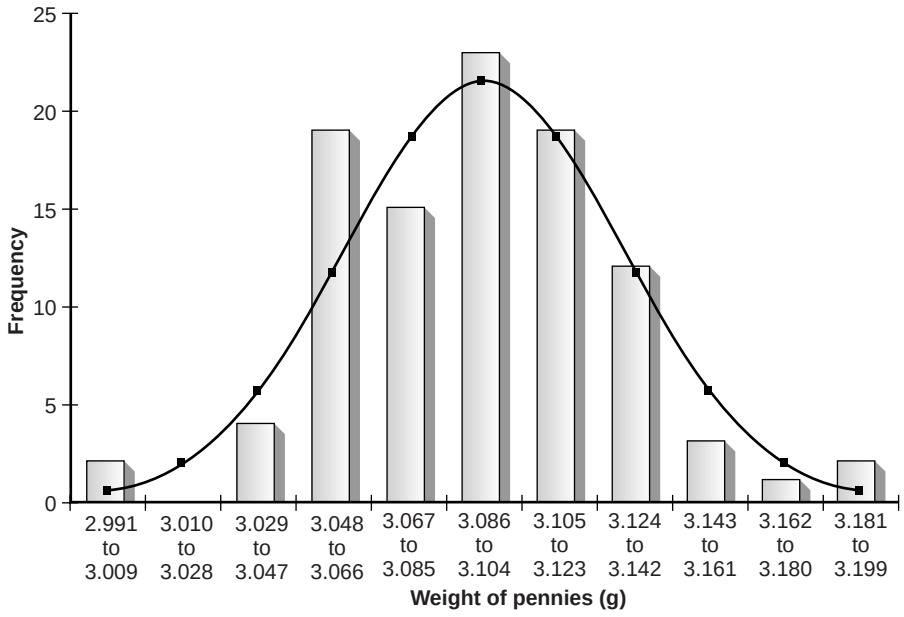


Figure 4.8
Histogram for data in Table 4.12. A normal distribution curve for the data, based on $\overline{X}$ and s^2, is superimposed on the histogram.

95%, and 100% of the pennies fall within these limits. It is easy to imagine that increasing the number of pennies in the sample will result in a histogram that even more closely approximates a normal distribution.

We will not offer a formal proof that the sample of pennies in Table 4.12 and the population from which they were drawn are normally distributed; however, the evidence we have seen strongly suggests that this is true. Although we cannot 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. That this is generally true is a consequence of the **central limit theorem.**[6] According to this theorem, in systems subject to a variety of indeterminate errors, the distribution of results will be approximately normal. Furthermore, as the number of contributing sources of indeterminate error increases, the results come even closer to approximating 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.

central limit theorem
The distribution of measurements subject to indeterminate errors is often a normal distribution.

Estimating μ and σ^2 Our comparison of the histogram for the data in Table 4.12 to a normal distribution assumes that the sample's mean, $\overline{X}$, and variance, s^2, are appropriate estimators of the population's mean, μ, and variance, σ^2. Why did we select $\overline{X}$ and s^2, as opposed to other possible measures of central tendency and spread? The explanation is simple; $\overline{X}$ and s^2 are considered unbiased estimators of μ and σ^2.[7,8] If we could analyze every possible sample of equal size for a given population (e.g., every possible sample of five pennies), calculating their respective means and variances, the average mean and the average variance would equal μ and σ^2. Although $\overline{X}$ and s^2 for any single sample probably will not be the same as μ or σ^2, they provide a reasonable estimate for these values.

Degrees of Freedom Unlike the population's variance, the variance of a sample includes the term $n - 1$ in the denominator, where n is the size of the sample

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n - 1} \qquad \textbf{4.12}$$

degrees of freedom
The number of independent values on which a result is based (ν).

Defining the sample's variance with a denominator of n, as in the case of the population's variance leads to a biased estimation of σ^2. The denominators of the variance equations 4.8 and 4.12 are commonly called the **degrees of freedom** for the population and the sample, respectively. In the case of a population, the degrees of freedom is always equal to the total number of members, n, in the population. For the sample's variance, however, substituting $\overline{X}$ for μ removes a degree of freedom from the calculation. That is, if there are n members in the sample, the value of the n^{th} member can always be deduced from the remaining $n - 1$ members and $\overline{X}$. For example, if we have a sample with five members, and we know that four of the members are 1, 2, 3, and 4, and that the mean is 3, then the fifth member of the sample must be

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

4D.5 Confidence Intervals for Samples

Earlier we introduced the confidence interval as a way to report the most probable value for a population's mean, μ, when the population's standard deviation, σ, is known. Since s^2 is an unbiased estimator of σ^2, it should be possible to construct confidence intervals for samples by replacing σ in equations 4.10 and 4.11 with s. Two complications arise, however. The first is that we cannot define s^2 for a single member of a population. Consequently, equation 4.10 cannot be extended to situations in which s^2 is used as an estimator of σ^2. In other words, when σ is unknown, we cannot construct a confidence interval for μ by sampling only a single member of the population.

The second complication is that the values of z shown in Table 4.11 are derived for a normal distribution curve that is a function of σ^2, not s^2. Although s^2 is an unbiased estimator of σ^2, the value of s^2 for any randomly selected sample may differ significantly from σ^2. To account for the uncertainty in estimating σ^2, the term z in equation 4.11 is replaced with the variable t, where t is defined such that $t \geq z$ at all confidence levels. Thus, equation 4.11 becomes

$$\mu = \overline{X} \pm \frac{ts}{\sqrt{n}} \qquad \textbf{4.13}$$

Values for t at the 95% confidence level are shown in Table 4.14. Note that t becomes smaller as the number of the samples (or degrees of freedom) increase, approaching z as n approaches infinity. Additional values of t for other confidence levels can be found in Appendix 1B.

Example

EXAMPLE 4.15

What is the 95% confidence interval for the data in Table 4.1?

SOLUTION

The mean and standard deviation for this sample are, respectively, 3.117 g and 0.051 g. Since the sample consists of seven measurements, there are six degrees

of freedom. The value of t from Table 4.14, is 2.45. Substituting into equation 4.13 gives

$$\mu = \overline{X} \pm \frac{ts}{\sqrt{n}} = 3.117 \pm \frac{(2.45)(0.051)}{\sqrt{7}} = 3.117 \pm 0.047 \text{ g}$$

Thus, there is a 95% probability that the population's mean is between 3.070 and 3.164 g.

Table 4.14 Values of *t* for the 95% Confidence Interval

Degrees of Freedom	*t*
1	12.71
2	4.30
3	3.18
4	2.78
5	2.57
6	2.45
7	2.36
8	2.31
9	2.26
10	2.23
12	2.18
14	2.14
16	2.12
18	2.10
20	2.09
30	2.04
50	2.01
∞	1.96

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 analyzing 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

$$\mu = 0.67 \pm \frac{(2.78)(0.64)}{\sqrt{5}} = 0.64 \pm 0.80 \text{ ppm}$$

This confidence interval states that the analyte's true concentration lies within the range of –0.16 ppm to 1.44 ppm. Including a negative concentration within the confidence interval should lead you to reevaluate your data or conclusions. On further investigation your data may show that the standard deviation is larger than expected,

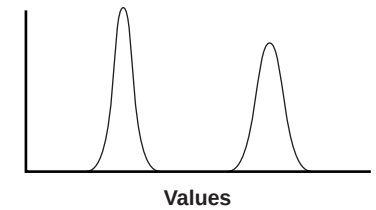
making the confidence interval too broad, or you may conclude that the analyte's concentration is too small to detect accurately.*

A second example is also informative. When samples are obtained from a normally distributed population, their values must be random. If results for several samples show a regular pattern or trend, then the samples cannot be normally distributed. This may reflect the fact that the underlying population is not normally distributed, or it may indicate the presence of a time-dependent determinate error. For example, if we randomly select 20 pennies and find that the mass of each penny exceeds that of the preceding penny, we might suspect that the balance on which the pennies are being weighed is drifting out of calibration.

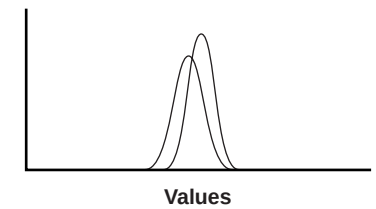
4E Statistical Analysis of Data

In the previous section we noted that the result of an analysis is best expressed as a confidence interval. For example, a 95% confidence interval for the mean of five results gives the range in which we expect to find the mean for 95% of all samples of equal size, drawn from the same population. Alternatively, and in the absence of determinate errors, the 95% confidence interval indicates the range of values in which we expect to find the population's true mean.

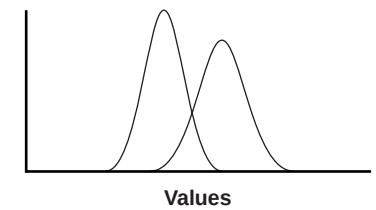
The probabilistic nature of a confidence interval provides an opportunity to ask and answer questions comparing a sample's mean or variance to either the accepted values for its population or similar values obtained for other samples. For example, confidence intervals can be used to answer questions such as "Does a newly developed method for the analysis of cholesterol in blood give results that are significantly different from those obtained when using a standard method?" or "Is there a significant variation in the chemical composition of rainwater collected at different sites downwind from a coalburning utility plant?" In this section we introduce a general approach to the statistical analysis of data. Specific statistical methods of analysis are covered in Section 4F.



(a)



(b)



(c)

Figure 4.9
Three examples of possible relationships between the probability distributions for two populations. (a) Completely separate distributions; (b) Distributions with a great deal of overlap; (c) Distributions with some overlap.

4E.1 Significance Testing

Let's consider the following problem. Two sets of blood samples have been collected from a patient receiving medication to lower her concentration of blood glucose. One set of samples was drawn immediately before the medication was administered; the second set was taken several hours later. The samples are analyzed and their respective means and variances reported. 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 probability distribution curves for each sample and to compare the curves with each other. Three possible outcomes are shown in Figure 4.9. In Figure 4.9a, the probability distribution curves are completely separated, strongly suggesting that the samples are significantly different. In Figure 4.9b, the probability distributions for the two samples are highly overlapped, suggesting that any difference between the samples is insignificant. Figure 4.9c, however, presents a dilemma. Although the means for the two samples appear to be different, the probability distributions overlap to an extent that a significant number of possible outcomes could belong to either distribution. In this case we can, at best, only make a statement about the probability that the samples are significantly different.

*The topic of detection limits is discussed at the end of this chapter.

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 turning to a discussion of specific examples, however, we will first establish a general approach to conducting and interpreting significance tests.

4E.2 Constructing a Significance Test

significance test
A statistical test to determine if the difference between two values is significant.

null hypothesis
A statement that the difference between two values can be explained by indeterminate error; retained if the significance test does not fail (H_0).

alternative hypothesis
A statement that the difference between two values is too great to be explained by indeterminate error; accepted if the significance test shows that null hypothesis should be rejected (H_A).

A **significance test** is designed 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 experimental problem as a yes-or-no question, two examples of which were given at the beginning of this section. 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 difference in the values being compared. The **alternative hypothesis,** H_A, is that the difference between the values is too great to be explained by random error and, therefore, must be real. A significance test is conducted on the null hypothesis, which is either retained or rejected. If the null hypothesis is rejected, then the alternative hypothesis must be accepted. When a null hypothesis is not rejected, it is said to be retained rather than accepted. A null hypothesis is retained whenever the evidence is insufficient to prove it is incorrect. Because of the way in which significance tests are conducted, it is impossible to prove that a null hypothesis is true.

The difference between retaining a null hypothesis and proving the null hypothesis is important. To appreciate this point, let us return to our example on determining the mass of a penny. After looking at the data in Table 4.12, you might pose the following null and alternative hypotheses

H_0: Any U.S. penny in circulation has a mass that falls in the range of 2.900–3.200 g

H_A: Some U.S. pennies in circulation have masses that are less than 2.900 g or more than 3.200 g.

To test the null hypothesis, you reach into your pocket, retrieve a penny, and determine its mass. If the mass of this penny is 2.512 g, then you have proved that the null hypothesis is incorrect. Finding that the mass of your penny is 3.162 g, however, does not prove that the null hypothesis is correct because the mass of the next penny you sample might fall outside the limits set by the null hypothesis.

After stating the null and alternative hypotheses, a significance level for the analysis is chosen. The significance level is the confidence level for retaining the null hypothesis or, in other words, the probability that the null hypothesis will be incorrectly rejected. In the former case the significance level is given as a percentage (e.g., 95%), whereas in the latter case, it is given as α, where α is defined as

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

Thus, for a 95% confidence level, α is 0.05.

Next, an equation for a test statistic is written, and the test statistic's critical value is found from an appropriate table. This critical value defines the breakpoint between values of the test statistic for which the null hypothesis will be retained or rejected. The test statistic is calculated from the data, compared with the critical value, and the null hypothesis is either rejected or retained. Finally, the result of the significance test is used to answer the original question.

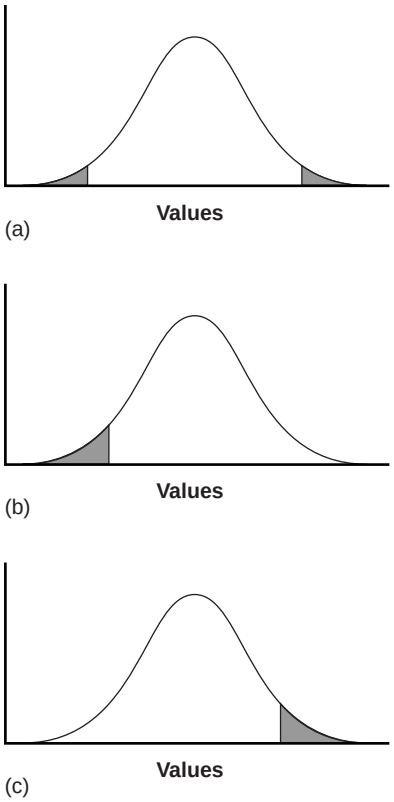


Figure 4.10
Examples of (a) two-tailed, (b) and (c) one-tailed, significance tests. The shaded areas in each curve represent the values for which the null hypothesis is rejected.

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

Consider the situation when the accuracy of a new analytical method is evaluated by analyzing a standard reference material with a known μ. A sample of the standard is analyzed, and the sample's mean is determined. The null hypothesis is that the sample's mean is equal to μ

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

If the significance test is conducted at the 95% confidence level ($\alpha = 0.05$), then the null hypothesis will be retained if a 95% confidence interval around $\overline{X}$ contains μ. If the alternative hypothesis is

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

then the null hypothesis will be rejected, and the alternative hypothesis accepted if μ lies in either of the shaded areas at the tails of the sample's probability distribution (Figure 4.10a). Each of the shaded areas accounts for 2.5% of the area under the probability distribution curve. This is called a **two-tailed significance test** because the null hypothesis is rejected for values of μ at either extreme of the sample's probability distribution.

The alternative hypothesis also can be stated in one of two additional ways

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

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

for which the null hypothesis is rejected if μ falls within the shaded areas shown in Figure 4.10(b) and Figure 4.10(c), respectively. In each case the shaded area represents 5% of the area under the probability distribution curve. These are examples of **one-tailed significance tests.**

For a fixed confidence level, a two-tailed test is always the more conservative test because it requires a larger difference between $\overline{X}$ and μ to reject the null hypothesis. Most significance tests are applied when there is no a priori expectation about the relative magnitudes of the parameters being compared. A two-tailed significance test, therefore, is usually the appropriate choice. One-tailed significance tests are reserved for situations when we have reason to expect one parameter to be larger or smaller than the other. For example, a one-tailed significance test would be appropriate for our earlier example regarding a medication's effect on blood glucose levels since we believe that the medication will lower the concentration of glucose.

two-tailed significance test
Significance test in which the null hypothesis is rejected for values at either end of the normal distribution.

one-tailed significance test
Significance test in which the null hypothesis is rejected for values at only one end of the normal distribution.

4E.4 Errors in Significance Testing

Since significance tests are based on probabilities, their interpretation is naturally subject to error. As we have already seen, significance tests are carried out at a significance level, α, that defines the probability of rejecting a null hypothesis that is true. For example, when a significance test is conducted at $\alpha = 0.05$, there is a 5% probability that the null hypothesis will be incorrectly rejected. This is known as a **type 1 error,** and its risk is always equivalent to α. Type 1 errors in two-tailed and one-tailed significance tests are represented by the shaded areas under the probability distribution curves in Figure 4.10.

The second type of error occurs when the null hypothesis is retained even though it is false and should be rejected. This is known as a **type 2 error,** and its probability of occurrence is β. Unfortunately, in most cases β cannot be easily calculated or estimated.

type 1 error
The risk of falsely rejecting the null hypothesis (α).

type 2 error
The risk of falsely retaining the null hypothesis (β).

The probability of a type 1 error is inversely related to the probability of a type 2 error. Minimizing a type 1 error by decreasing α, for example, increases the likelihood of a type 2 error. The value of α chosen for a particular significance test, therefore, represents a compromise between these two types of error. Most of the examples in this text use a 95% confidence level, or $\alpha = 0.05$, since this is the most frequently used confidence level for the majority of analytical work. It is not unusual, however, for more stringent (e.g. $\alpha = 0.01$) or for more lenient (e.g. $\alpha = 0.10$) confidence levels to be used.

4F Statistical Methods for Normal Distributions

The most commonly encountered probability distribution is the normal, or Gaussian, distribution. A normal distribution is characterized by a true mean, μ, and variance, σ^2, which are estimated using $\bar{X}$ and s^2. Since the area between any two limits of a normal distribution is well defined, the construction and evaluation of significance tests are straightforward.

4F.1 Comparing $\bar{X}$ to μ

One approach for validating a new analytical method is to analyze a standard sample containing a known amount of analyte, μ. The method's accuracy is judged by determining the average amount of analyte in several samples, $\bar{X}$, and using a significance test to compare it with μ. The null hypothesis is that $\bar{X}$ and μ are the same and that any difference between the two values can be explained by indeterminate errors affecting the determination of $\bar{X}$. The alternative hypothesis is that the difference between $\bar{X}$ and μ is too large to be explained by indeterminate error.

The equation for the test (experimental) statistic, t_{exp}, is derived from the confidence interval for μ

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

Rearranging equation 4.14

$$t_{exp} = \frac{|\mu - \bar{X}| \times \sqrt{n}}{s} \qquad \textbf{4.15}$$

gives the value of t_{exp} when μ is at either the right or left edge of the sample's apparent confidence interval (Figure 4.11a). The value of t_{exp} is compared with a critical value, $t(\alpha,\nu)$, which is determined by the chosen significance level, α, the degrees of freedom for the sample, ν, and whether the significance test is one-tailed or two-tailed. Values for $t(\alpha,\nu)$ are found in Appendix 1B. The critical value $t(\alpha,\nu)$ defines the confidence interval that can be explained by indeterminate errors. If t_{exp} is greater than $t(\alpha,\nu)$, then the confidence interval for the data is wider than that expected from indeterminate errors (Figure 4.11b). In this case, the null hypothesis is rejected and the alternative hypothesis is accepted. If t_{exp} is less than or equal to $t(\alpha,\nu)$, then the confidence interval for the data could be attributed to indeterminate error, and the null hypothesis is retained at the stated significance level (Figure 4.11c).

A typical application of this significance test, which is known as a ***t*-test** of $\bar{X}$ to μ, is outlined in the following example.

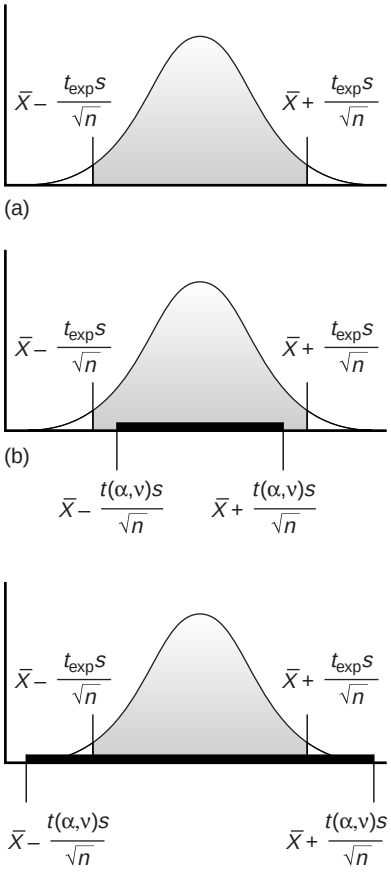


Figure 4.11
Relationship between confidence intervals and results of a significance test. (a) The shaded area under the normal distribution curves shows the apparent confidence intervals for the sample based on t_{exp}. The solid bars in (b) and (c) show the actual confidence intervals that can be explained by indeterminate error using the critical value of (α,ν). In part (b) the null hypothesis is rejected and the alternative hypothesis is accepted. In part (c) the null hypothesis is retained.

***t*-test**
Statistical test for comparing two mean values to see if their difference is too large to be explained by indeterminate error.

Example

EXAMPLE 4.16

Before determining the amount of Na_2CO_3 in an unknown sample, a student decides to check her procedure by analyzing a sample known to contain 98.76% w/w Na_2CO_3. Five replicate determinations of the %w/w Na_2CO_3 in the standard were made with the following results

98.71% 98.59% 98.62% 98.44% 98.58%

Is the mean for these five trials significantly different from the accepted value at the 95% confidence level ($\alpha = 0.05$)?

SOLUTION

The mean and standard deviation for the five trials are

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

Since there is no reason to believe that $\overline{X}$ must be either larger or smaller than μ, the use of a two-tailed significance test is appropriate. The null and alternative hypotheses are

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

The test statistic is

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

The critical value for $t(0.05,4)$, as found in Appendix 1B, is 2.78. Since t_{exp} is greater than $t(0.05, 4)$, we must reject the null hypothesis and accept the alternative hypothesis. At the 95% confidence level the difference between $\overline{X}$ and μ is significant and cannot be explained by indeterminate sources of error. There is evidence, therefore, that the results are affected by a determinate source of error.

If evidence for a determinate error is found, as in Example 4.16, its source should be identified and corrected before analyzing additional samples. Failing to reject the null hypothesis, however, does not imply that the method is accurate, but only indicates that there is insufficient evidence to prove the method inaccurate at the stated confidence level.

The utility of the *t*-test for $\overline{X}$ and μ is improved by optimizing the conditions used in determining $\overline{X}$. Examining equation 4.15 shows that increasing the number of replicate determinations, *n,* or improving the precision of the analysis enhances the utility of this significance test. A *t*-test can only give useful results, however, if the standard deviation for the analysis is reasonable. If the standard deviation is substantially larger than the expected standard deviation, σ, the confidence interval around $\overline{X}$ will be so large that a significant difference between $\overline{X}$ and μ may be difficult to prove. On the other hand, if the standard deviation is significantly smaller than expected, the confidence interval around $\overline{X}$ will be too small, and a significant difference between $\overline{X}$ and μ may be found when none exists. A significance test that can be used to evaluate the standard deviation is the subject of the next section.

4F.2 Comparing s^2 to σ^2

When a particular type of sample is analyzed on a regular basis, it may be possible to determine the expected, or true variance, σ^2, for the analysis. This often is the case in clinical labs where hundreds of blood samples are analyzed each day. Replicate analyses of any single sample, however, results in a sample variance, s^2. A statistical comparison of s^2 to σ^2 provides useful information about whether the analysis is in a state of "statistical control." The null hypothesis is that s^2 and σ^2 are identical, and the alternative hypothesis is that they are not identical.

The test statistic for evaluating the null hypothesis is called an ***F*-test,** and is given as either

$$F_{exp} = \frac{s^2}{\sigma^2} \quad (s^2 > \sigma^2) \qquad \text{or} \qquad F_{exp} = \frac{\sigma^2}{s^2} \quad (\sigma^2 > s^2) \tag{4.16}$$

depending on whether s^2 is larger or smaller than σ^2. Note that F_{exp} is defined such that its value is always greater than or equal to 1.

***F*-test**
Statistical test for comparing two variances to see if their difference is too large to be explained by indeterminate error.

If the null hypothesis is true, then F_{exp} should equal 1. Due to indeterminate errors, however, the value for F_{exp} usually is greater than 1. A critical value, $F(\alpha, \nu_{num}, \nu_{den})$, gives the largest value of F that can be explained by indeterminate error. It is chosen for a specified significance level, α, and the degrees of freedom for the variances in the numerator, ν_{num}, and denominator, ν_{den}. The degrees of freedom for s^2 is $n - 1$, where n is the number of replicates used in determining the sample's variance. Critical values of F for $\alpha = 0.05$ are listed in Appendix 1C for both one-tailed and two-tailed significance tests.

Example

EXAMPLE 4.17

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

254 249 252 252 249 249 250 247 251 252

Determine whether there is any evidence that the measurement process is not under statistical control at $\alpha = 0.05$.

SOLUTION

The variance for the sample of ten tablets is 4.3. A two-tailed significance test is used since the measurement process is considered out of statistical control if the sample's variance is either too good or too poor. The null hypothesis and alternative hypotheses are

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

The test statistic is

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

The critical value for $F(0.05, \infty, 9)$ from Appendix 1C is 3.33. Since F is greater than $F(0.05,\infty, 9)$, we reject the null hypothesis and accept the alternative hypothesis that the analysis is not under statistical control. One explanation for the unreasonably small variance could be that the aspirin tablets were not selected randomly.

4F.3 Comparing Two Sample Variances

The *F*-test can be extended to the comparison of variances for two samples, A and B, by rewriting equation 4.16 as

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

where A and B are defined such that s_A^2 is greater than or equal to s_B^2. An example of this application of the *F*-test is shown in the following example.

Example

EXAMPLE 4.18

Tables 4.1 and 4.8 show results for two separate 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

Letting A represent the results in Table 4.1 and B represent the results in Table 4.8, we find that the variances are $s_A^2 = 0.00259$ and $s_B^2 = 0.00138$. A two-tailed significance test is used since there is no reason to suspect that the results for one analysis will be more precise than that of the other. The null and alternative hypotheses are

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

and the test statistic is

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

The critical value for $F(0.05, 6, 4)$ is 9.197. Since F_{exp} is less than $F(0.05, 6, 4)$, the null hypothesis is retained. There is no evidence at the chosen significance level to suggest that the difference in precisions is significant.

4F.4 Comparing Two Sample Means

The result of an analysis is influenced by three factors: the method, the sample, and the analyst. The influence of these factors can be studied by conducting a pair of experiments in which only one factor is changed. For example, two methods can be compared by having the same analyst apply both methods to the same sample and examining the resulting means. In a similar fashion, it is possible to compare two analysts or two samples.

Significance testing for comparing two mean values is divided into two categories depending on the source of the data. Data are said to be **unpaired** when each mean is derived from the analysis of several samples drawn from the same source. **Paired data** are encountered when analyzing a series of samples drawn from different sources.

unpaired data
Two sets of data consisting of results obtained using several samples drawn from a single source.

paired data
Two sets of data consisting of results obtained using several samples drawn from different sources.

Unpaired Data Consider two samples, A and B, for which mean values, $\overline{X}_A$ and $\overline{X}_B$, and standard deviations, s_A and s_B, have been measured. Confidence intervals for μ_A and μ_B can be written for both samples

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

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

where n_A and n_B are the number of replicate trials conducted on samples A and B. A comparison of the mean values is based on the null hypothesis that $\overline{X}_A$ and $\overline{X}_B$ are identical, and an alternative hypothesis that the means are significantly different.

A test statistic is derived by letting μ_A equal μ_B, and combining equations 4.17 and 4.18 to give

$$\overline{X}_A \pm \frac{ts_A}{\sqrt{n_A}} = \overline{X}_B \pm \frac{ts_B}{\sqrt{n_B}}$$

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

$$\left|\overline{X}_A - \overline{X}_B\right| = t \times \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

Finally, solving for t, which we replace with t_{exp}, leaves us with

$$t_{exp} = \frac{\left|\overline{X}_A - \overline{X}_B\right|}{\sqrt{(s_A^2/n_A) + (s_B^2/n_B)}} \quad \textbf{4.19}$$

The value of t_{exp} is compared with a critical value, $t(\alpha, \nu)$, as determined by the chosen significance level, α, the degrees of freedom for the sample, ν, and whether the significance test is one-tailed or two-tailed.

It is unclear, however, how many degrees of freedom are associated with $t(\alpha, \nu)$ since there are two sets of independent measurements. If the variances s_A^2 and s_B^2 estimate the same σ^2, then the two standard deviations can be factored out of equation 4.19 and replaced by a pooled standard deviation, s_{pool}, which provides a better estimate for the precision of the analysis. Thus, equation 4.19 becomes

$$t_{exp} = \frac{\left|\overline{X}_A - \overline{X}_B\right|}{s_{pool}\sqrt{(1/n_A) + (1/n_B)}} \quad \textbf{4.20}$$

with the pooled standard deviation given as

$$s_{pool} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}} \quad \textbf{4.21}$$

As indicated by the denominator of equation 4.21, the degrees of freedom for the pooled standard deviation is $n_A + n_B - 2$.

If s_A and s_B are significantly different, however, then t_{exp} must be calculated using equation 4.19. In this case, the degrees of freedom is calculated using the following imposing equation.

$$\nu = \frac{[(s_A^2/n_A) + (s_B^2/n_B)]^2}{[(s_A^2/n_A)^2/(n_A + 1)] + [(s_B^2/n_B)^2/(n_B + 1)]} - 2 \quad \textbf{4.22}$$

Since the degrees of freedom must be an integer, the value of ν obtained using equation 4.22 is rounded to the nearest integer.

Regardless of whether equation 4.19 or 4.20 is used to calculate t_{exp}, the null hypothesis is rejected if t_{exp} is greater than $t(\alpha, \nu)$, and retained if t_{exp} is less than or equal to $t(\alpha, \nu)$.

Example

EXAMPLE 4.19

Tables 4.1 and 4.8 show results for two separate 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

To begin with, we must determine whether the variances for the two analyses are significantly different. This is done using an *F*-test as outlined in Example 4.18. Since no significant difference was found, a pooled standard deviation with 10 degrees of freedom is calculated

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

$$= \sqrt{\frac{(7 - 1)(0.00259) + (5 - 1)(0.00138)}{7 + 5 - 2}}$$

$$= 0.0459$$

where the subscript A indicates the data in Table 4.1, and the subscript B indicates the data in Table 4.8. The comparison of the means for the two analyses is based on the null hypothesis

$$H_0: \quad \overline{X}_A = \overline{X}_B$$

and a two-tailed alternative hypothesis

$$H_A: \quad \overline{X}_A \neq \overline{X}_B$$

Since the standard deviations can be pooled, the test statistic is calculated using equation 4.20

$$t_{exp} = \frac{\left|\overline{X}_A - \overline{X}_B\right|}{s_{pool}\sqrt{(1/n_A + 1/n_B)}} = \frac{\left|3.117 - 3.081\right|}{0.0459\sqrt{(1/7 + 1/5)}} = 1.34$$

The critical value for $t(0.05, 10)$, from Appendix 1B, is 2.23. Since t_{exp} is less than $t(0.05, 10)$ the null hypothesis is retained, and there is no evidence that the two sets of pennies are significantly different at the chosen significance level.

Example

EXAMPLE 4.20

The %w/w Na_2CO_3 in soda ash can be determined by an acid–base titration. The results obtained by two analysts are shown here. Determine whether the difference in their mean values is significant at $\alpha = 0.05$.

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

SOLUTION

We begin by summarizing the mean and standard deviation for the data reported by each analyst. These values are

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

A two-tailed *F*-test of the following null and alternative hypotheses

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

is used to determine whether a pooled standard deviation can be calculated. The test statistic is

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

Since F_{exp} is larger than the critical value of 7.15 for $F(0.05, 5, 5)$, the null hypothesis is rejected and the alternative hypothesis that the variances are significantly different is accepted. As a result, a pooled standard deviation cannot be calculated.

The mean values obtained by the two analysts are compared using a two-tailed *t*-test. The null and alternative hypotheses are

$$H_0: \quad \overline{X}_A = \overline{X}_B \qquad H_{A:} \quad \overline{X}_A \neq \overline{X}_B$$

Since a pooled standard deviation could not be calculated, the test statistic, t_{exp}, is calculated using equation 4.19

$$t_{exp} = \frac{\left|\overline{X}_A - \overline{X}_B\right|}{\sqrt{(s_A^2/n_A) + (s_B^2/n_B)}} = \frac{\left|86.83 - 82.71\right|}{\sqrt{[(0.32)^2/6] + [(2.16)^2/6]}} = 4.62$$

and the degrees of freedom are calculated using equation 4.22

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

The critical value for $t(0.05, 5)$ is 2.57. Since the calculated value of t_{exp} is greater than $t(0.05, 5)$ we reject the null hypothesis and accept the alternative hypothesis that the mean values for %w/w Na_2CO_3 reported by the two analysts are significantly different at the chosen significance level.

Paired Data In some situations the variation within the data sets being compared is more significant than the difference between the means of the two data sets. This is commonly encountered in clinical and environmental studies, where the data being compared usually consist of a set of samples drawn from several populations. For example, a study designed to investigate two procedures for monitoring the concentration of glucose in blood might involve blood samples drawn from ten patients. If the variation in the blood glucose levels among the patients is significantly larger than the anticipated variation between the methods, then an analysis in which the data are treated as unpaired will fail to find a significant difference between the

methods. In general, paired data sets are used whenever the variation being investigated is smaller than other potential sources of variation.

In a study involving paired data the difference, d_i, between the paired values for each sample is calculated. The average difference, $\bar{d}$, and standard deviation of the differences, s_d, are then calculated. The null hypothesis is that $\bar{d}$ is 0, and that there is no difference in the results for the two data sets. The alternative hypothesis is that the results for the two sets of data are significantly different, and, therefore, $\bar{d}$ is not equal to 0.

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

$$0 = \bar{d} \pm \frac{ts_d}{\sqrt{n}}$$

where n is the number of paired samples. Replacing t with t_{exp} and rearranging gives

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

The value of t_{exp} is then compared with a critical value, $t(\alpha, \nu)$, which is determined by the chosen significance level, α, the degrees of freedom for the sample, ν, and whether the significance test is one-tailed or two-tailed. For paired data, the degrees of freedom is $n - 1$. If t_{exp} is greater than $t(\alpha, \nu)$, then the null hypothesis is rejected and the alternative hypothesis is accepted. If t_{exp} is less than or equal to $t(\alpha, \nu)$, then the null hypothesis is retained, and a significant difference has not been demonstrated at the stated significance level. This is known as the **paired *t*-test.**

paired *t*-test
Statistical test for comparing paired data to determine if their difference is too large to be explained by indeterminate error.

Example

EXAMPLE 4.21

Marecek and colleagues developed a new electrochemical method for the rapid quantitative analysis of the antibiotic monensin in the fermentation vats used during its production.[9] The standard method for the analysis, which is based on a test for microbiological activity, is both difficult and time-consuming. As part of the study, samples taken at different times from a fermentation production vat were analyzed for the concentration of monensin using both the electrochemical and microbiological procedures. 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

Determine whether there is a significant difference between the methods at $\alpha = 0.05$.

*1 ppt is equivalent to 0.1%.

SOLUTION

This is an example of a paired data set since the acquisition of samples over an extended period introduces a substantial time-dependent change in the concentration of monensin. The comparison of the two methods must be done with the paired *t*-test, using the following null and two-tailed alternative hypotheses

$$H_0: \quad \bar{d} = 0 \qquad H_{A:} \quad \bar{d} \neq 0$$

Defining the difference between the methods as

$$d = X_{\text{elect}} - X_{\text{micro}}$$

we can calculate the difference for each sample

Sample	1	2	3	4	5	6	7	8	9	10	11
d	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 2.25 and 5.63, respectively. The test statistic is

$$t_{\text{exp}} = \frac{|\bar{d}|\sqrt{n}}{s_d} = \frac{|2.25|\sqrt{11}}{5.63} = 1.33$$

which is smaller than the critical value of 2.23 for $t(0.05, 10)$. Thus, the null hypothesis is retained, and there is no evidence that the two methods yield different results at the stated significance level.

A paired *t*-test can only be applied when the individual differences, d_i, belong to the same population. This will only be true if the determinate and indeterminate errors affecting the results are independent of the concentration of analyte in the samples. If this is not the case, a single sample with a larger error could result in a value of d_i that is substantially larger than that for the remaining samples. Including this sample in the calculation of $\bar{d}$ and s_d leads to a biased estimate of the true mean and standard deviation. For samples that span a limited range of analyte concentrations, such as that in Example 4.21, this is rarely a problem. 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 the 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 comparison is best made using a linear regression, details of which are discussed in the next chapter.

4F.5 Outliers

On occasion, a data set appears to be skewed by the presence of one or more data points that are not consistent with the remaining data points. Such values are called **outliers.** The most commonly used significance test for identifying outliers is **Dixon's *Q*-test.** The null hypothesis is that the apparent outlier is taken from the same population as the remaining data. The alternative hypothesis is that the outlier comes from a different population, and, therefore, should be excluded from consideration.

outlier
Data point whose value is much larger or smaller than the remaining data.

Dixon's *Q*-test
Statistical test for deciding if an outlier can be removed from a set of data.

The *Q*-test compares the difference between the suspected outlier and its nearest numerical neighbor to the range of the entire data set. Data are ranked from smallest to largest so that the suspected outlier is either the first or the last data

point. The test statistic, Q_{exp}, is calculated using equation 4.23 if the suspected outlier is the smallest value (X_1)

$$Q_{exp} = \frac{X_2 - X_1}{X_n - X_1} \qquad \textbf{4.23}$$

or using equation 4.24 if the suspected outlier is the largest value (X_n)

$$Q_{exp} = \frac{X_n - X_{n-1}}{X_n - X_1} \qquad \textbf{4.24}$$

where n is the number of members in the data set, including the suspected outlier. It is important to note that equations 4.23 and 4.24 are valid only for the detection of a single outlier. Other forms of Dixon's Q-test allow its extension to the detection of multiple outliers.[10] The value of Q_{exp} is compared with a critical value, $Q(\alpha, n)$, at a significance level of α. The Q-test is usually applied as the more conservative two-tailed test, even though the outlier is the smallest or largest value in the data set. Values for $Q(\alpha, n)$ can be found in Appendix 1D. If Q_{exp} is greater than $Q(\alpha, n)$, then the null hypothesis is rejected and the outlier may be rejected. When Q_{exp} is less than or equal to $Q(\alpha, n)$ the suspected outlier must be retained.

Example

EXAMPLE 4.22

The following masses, in grams, were recorded in an experiment to determine the average mass of a U.S. penny.

3.067 3.049 3.039 2.514 3.048 3.079 3.094 3.109 3.102

Determine if the value of 2.514 g is an outlier at $\alpha = 0.05$.

SOLUTION

To begin with, place the masses in order from smallest to largest

2.514 3.039 3.048 3.049 3.067 3.079 3.094 3.102 3.109

and calculate Q_{exp}

$$Q_{exp} = \frac{X_2 - X_1}{X_9 - X_1} = \frac{3.039 - 2.514}{3.109 - 2.514} = 0.882$$

The critical value for $Q(0.05, 9)$ is 0.493. Since $Q_{exp} > Q(0.05, 9)$ the value is assumed to be an outlier, and can be rejected.

The Q-test should be applied with caution since there is a probability, equal to α, that an outlier identified by the Q-test actually is not an outlier. In addition, the Q-test should be avoided when rejecting an outlier leads to a precision that is unreasonably better than the expected precision determined by a propagation of uncertainty. Given these two concerns it is not surprising that some statisticians caution against the removal of outliers.[11] On the other hand, testing for outliers can provide useful information if we try to understand the source of the suspected outlier. For example, the outlier identified in Example 4.22 represents a significant change in the mass of a penny (an approximately 17% decrease in mass), due to 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.[12] The pennies in Example 4.22 were therefore drawn from different populations.

4G Detection Limits

The focus of this chapter has been the evaluation of analytical data, including the use of statistics. In this final section we consider how statistics may be used to characterize a method's ability to detect trace amounts of an analyte.

A method's **detection limit** is the smallest amount or concentration of analyte that can be detected with statistical confidence. The International Union of Pure and Applied Chemistry (IUPAC) defines the detection limit as the smallest concentration or absolute amount of analyte that has a signal significantly larger than the signal arising from a reagent blank. Mathematically, the analyte's signal at the detection limit, $(S_A)_{DL}$, is

$$(S_A)_{DL} = S_{reag} + z\sigma_{reag} \qquad \textbf{4.25}$$

where S_{reag} is the signal for a reagent blank, σ_{reag} is the known standard deviation for the reagent blank's signal, and z is a factor accounting for the desired confidence level. The concentration, $(C_A)_{DL}$, or absolute amount of analyte, $(n_A)_{DL}$, at the detection limit can be determined from the signal at the detection limit.

$$(C_A)_{DL} = \frac{(S_A)_{DL}}{k}$$

$$(n_A)_{DL} = \frac{(S_A)_{DL}}{k}$$

The value for z depends on the desired significance level for reporting the detection limit. Typically, z is set to 3, which, from Appendix 1A, corresponds to a significance level of $\alpha = 0.00135$. Consequently, only 0.135% of measurements made on the blank will yield signals that fall outside this range (Figure 4.12a). When σ_{reag} is unknown, the term $z\sigma_{reag}$ may be replaced with ts_{reag}, where t is the appropriate value from a t-table for a one-tailed analysis.[13]

In analyzing a sample to determine whether an analyte is present, the signal for the sample is compared with the signal for the blank. The null hypothesis is that the sample does not contain any analyte, in which case $(S_A)_{DL}$ and S_{reag} are identical. The alternative hypothesis is that the analyte is present, and $(S_A)_{DL}$ is greater than S_{reag}. If $(S_A)_{DL}$ exceeds S_{reag} by $z\sigma$(or ts), then the null hypothesis is rejected and there is evidence for the analyte's presence in the sample. The probability that the null hypothesis will be falsely rejected, a type 1 error, is the same as the significance level. Selecting z to be 3 minimizes the probability of a type 1 error to 0.135%.

Significance tests, however, also are subject to type 2 errors in which the null hypothesis is falsely retained. Consider, for example, the situation shown in Figure 4.12b, where S_A is exactly equal to $(S_A)_{DL}$. In this case the probability of a type 2 error is 50% since half of the signals arising from the sample's population fall below the detection limit. Thus, there is only a 50:50 probability that an analyte at the IUPAC detection limit will be detected. As defined, the IUPAC definition for the detection limit only indicates the smallest signal for which we can say, at a significance level of α, that an analyte is present in the sample. Failing to detect the analyte, however, does not imply that it is not present.

An alternative expression for the detection limit, which minimizes both type 1 and type 2 errors, is the **limit of identification,** $(S_A)_{LOI}$, which is defined as [14]

$$(S_A)_{LOI} = S_{reag} + z\sigma_{reag} + z\sigma_{samp}$$

detection limit
The smallest concentration or absolute amount of analyte that can be reliably detected.

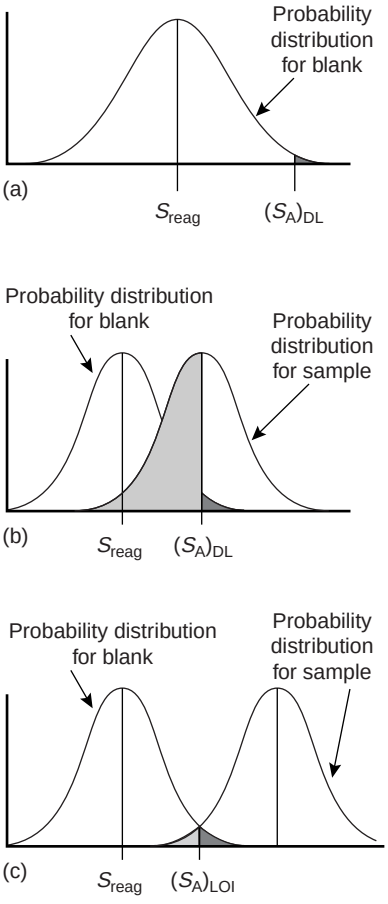


Figure 4.12
Normal distribution curves showing the definition of detection limit and limit of identification (LOI). The probability of a type 1 error is indicated by the dark shading, and the probability of a type 2 error is indicated by light shading.

limit of identification
The smallest concentration or absolute amount of analyte such that the probability of type 1 and type 2 errors are equal (LOI).

limit of quantitation
The smallest concentration or absolute amount of analyte that can be reliably determined (LOQ).

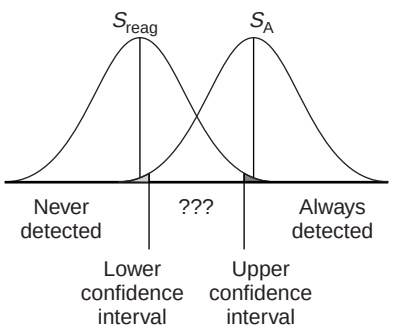


Figure 4.13
Establishment of areas where the signal is never detected, always detected, and where results are ambiguous. The upper and lower confidence limits are defined by the probability of a type 1 error (*dark shading*), and the probability of a type 2 error (*light shading*).

As shown in Figure 4.12c, the limit of identification is selected such that there is an equal probability of type 1 and type 2 errors. The American Chemical Society's Committee on Environmental Analytical Chemistry recommends the **limit of quantitation,** $(S_A)_{LOQ}$, which is defined as[15]

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

Other approaches for defining the detection limit have also been developed.[16]

The detection limit is often represented, particularly when used in debates over public policy issues, as a distinct line separating analytes that can be detected from those that cannot be detected.[17] This use of a detection limit is incorrect. Defining the detection limit in terms of statistical confidence levels implies that there may be a gray area where the analyte is sometimes detected and sometimes not detected. This is shown in Figure 4.13 where the upper and lower confidence limits are defined by the acceptable probabilities for type 1 and type 2 errors. Analytes producing signals greater than that defined by the upper confidence limit are always detected, and analytes giving signals smaller than the lower confidence limit are never detected. Signals falling between the upper and lower confidence limits, however, are ambiguous because they could belong to populations representing either the reagent blank or the analyte. Figure 4.12c represents the smallest value of S_A for which no such ambiguity exists.

4H KEY TERMS

alternative hypothesis (*p. 83*)
binomial distribution (*p. 72*)
central limit theorem (*p. 79*)
confidence interval (*p. 75*)
constant determinate error (*p. 60*)
degrees of freedom (*p. 80*)
detection limit (*p. 95*)
determinate error (*p. 58*)
Dixon's *Q*-test (*p. 93*)
error (*p. 64*)
F-test (*p. 87*)
heterogeneous (*p. 58*)
histogram (*p. 77*)
homogeneous (*p. 72*)
indeterminate error (*p. 62*)
limit of identification (*p. 95*)
limit of quantitation (*p. 96*)
mean (*p. 54*)
measurement error (*p. 58*)
median (*p. 55*)
method error (*p. 58*)
normal distribution (*p. 73*)
null hypothesis (*p. 83*)
one-tailed significance test (*p. 84*)
outlier (*p. 93*)
paired data (*p. 88*)
paired *t*-test (*p. 92*)
personal error (*p. 60*)
population (*p. 71*)
probability distribution (*p. 71*)
proportional determinate error (*p. 61*)
range (*p. 56*)
repeatability (*p. 62*)
reproducibility (*p. 62*)
sample (*p. 71*)
sampling error (*p. 58*)
significance test (*p. 83*)
standard deviation (*p. 56*)
standard reference material (*p. 61*)
tolerance (*p. 58*)
t-test (*p. 85*)
two-tailed significance test (*p. 84*)
type 1 error (*p. 84*)
type 2 error (*p. 84*)
uncertainty (*p. 64*)
unpaired data (*p. 88*)
variance (*p. 57*)

4I SUMMARY

The data we collect are characterized by their central tendency (where the values are clustered), and their spread (the variation of individual values around the central value). Central tendency is reported by stating the mean or median. The range, standard deviation, or variance may be used to report the data's spread. Data also are characterized by their errors, which include determinate errors affecting the data's accuracy, and indeterminate errors affecting the data's precision. A propagation of uncertainty allows us to estimate the affect of these determinate and indeterminate errors on results determined from our data.

The distribution of the results of an analysis around a central value is often 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 results within a specified range. For a normal distribution this range is best expressed 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. A *t*-test is used to compare mean values, and an *F*-test to compare precisions. Comparisons between two sets of data require an initial evaluation of whether the data is paired or unpaired. For unpaired data it is also necessary to 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.

Finally, we have seen that the detection limit is a statistical statement about the smallest amount of analyte that can be detected with confidence. A detection limit is not exact because 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.

4J *Suggested* EXPERIMENTS

Experiments

The following experiments may be used to introduce the statistical analysis of data in the analytical chemistry laboratory. Each experiment is annotated with a brief description of the data collected and the type of statistical analysis used in evaluating the data.

Cunningham, C. C.; Brown, G. R.; St Pierre, L. E. "Evaluation of Experimental Data," *J. Chem. Educ.* **1981,** *58,* 509–511.

In this experiment students determine the density of glass marbles and the radius of the bore of a glass capillary tube. Density is determined by measuring a marble's mass and volume, the latter by measuring a marble's diameter and assuming a spherical shape. Results are compared with those expected for a normal distribution. The radius of a glass capillary tube is determined using Poiseuille's equation by measuring the volume flow rate of water as a function of the hydrostatic head. In both experiments the experimentally obtained standard deviation is compared with that estimated by a propagation of uncertainty.

Gordus, A. A. "Statistical Evaluation of Class Data for Two Buret Readings," *J. Chem. Educ.* **1987,** *64,* 376–377.

The volumes of water in two burets are read, and the difference between the volumes are calculated. Students analyze the data by drawing histograms for each of the three volumes, comparing results with those predicted for a normal distribution.

Harvey, D. T. "Statistical Evaluation of Acid/Base Indicators," *J. Chem. Educ.* **1991,** *68,* 329–331.

In this experiment students standardize a solution of HCl by titration using several different indicators to signal the titration's end point. A statistical analysis of the data using *t*-tests and *F*-tests allows students to compare results obtained using the same indicator, with results obtained using different indicators. The results of this experiment can be used later when discussing the selection of appropriate indicators.

O'Reilley, J. E. "The Length of a Pestle," *J. Chem. Educ.* **1986,** *63,* 894–896.

In this experiment students measure the length of a pestle using a wooden meter stick, a stainless-steel ruler, and a vernier caliper. The data collected in this experiment provide an opportunity to discuss significant figures and sources of error. Statistical analysis includes the *Q*-test, *t*-test, and *F*-test.

Paselk, R. A. "An Experiment for Introducing Statistics to Students of Analytical and Clinical Chemistry," *J. Chem. Educ.* **1985,** *62,* 536.

Students use a commercial diluter to prepare five sets of dilutions of a stock dye solution (each set contains ten replicates) using two different diluters. Results are compared using *t*-tests and *F*-tests.

Richardson, T. H. "Reproducible Bad Data for Instruction in Statistical Methods," *J. Chem. Educ.* **1991,** *68,* 310–311.

This experiment uses the change in the mass of a U.S. penny to create data sets with outliers. Students are given a sample of ten pennies, nine of which are from one population. The *Q*-test is used to verify that the outlier can be rejected. Class data from each of the two populations of pennies are pooled and compared with results predicted for a normal distribution.

Sheeran, D. "Copper Content in Synthetic Copper Carbonate: A Statistical Comparison of Experimental and Expected Results," *J. Chem. Educ.* **1998,** *75,* 453–456.

In this experiment students synthesize basic copper(II) carbonate and determine the %w/w Cu by reducing the copper to Cu. A statistical analysis of the results shows that the synthesis does not produce $CuCO_3$, the compound that many predict to be the product (although it does not exist). Results are shown to be consistent with a hemihydrate of malachite, $Cu_2(OH)_2(CO_3) \cdot 1/2H_2O$, or azurite, $Cu_3(OH)_2(CO_3)_2$.

—Continued

Experiments

Continued from page 97

Spencer, R. D. "The Dependence of Strength in Plastics upon Polymer Chain Length and Chain Orientation," *J. Chem. Educ.* **1984,** *61,* 555–563.

The stretching properties of polymers are investigated by examining the effect of polymer orientation, polymer chain length, stretching rate, and temperature. Homogeneity of polymer films and consistency between lots of polymer films also are investigated. Statistical analysis of data includes *Q*-tests and *t*-tests.

Thomasson, K.; Lofthus-Merschman, S.; Humbert, M.; et al. "Applying Statistics in the Undergraduate Chemistry Laboratory: Experiments with Food Dyes," *J. Chem. Educ.* **1998,** *75,* 231–233.

The absorbance of solutions of food dyes is used to explore the treatment of outliers and the application of the *t*-test for comparing means.

Vitha, M. F.; Carr, P. W. "A Laboratory Exercise in Statistical Analysis of Data," *J. Chem. Educ.* **1997,** *74,* 998–1000.

Students determine the average weight of vitamin E pills using several different methods (one at a time, in sets of ten pills, and in sets of 100 pills). The data collected by the class are pooled together, plotted as histograms, and compared with results predicted by a normal distribution. The histograms and standard deviations for the pooled data also show the effect of sample size on the standard error of the mean.

4K PROBLEMS

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

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

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

2. Shown in the following rows are results for the determination of acetaminophen (in milligrams) in ten separate tablets of Excedrin Extra Strength Pain Reliever.[18]

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 these data. (b) Assuming that $\bar{X}$ and s^2 are good approximations for μ and σ^2, and that the population is normally distributed, what percentage of tablets are expected to contain more than the standard amount of 250 mg acetaminophen per tablet?

3. Salem and Galan have developed a new method for determining the amount of morphine hydrochloride in tablets.[19] Results, in milligrams, for several tablets containing different nominal dosages follow

100-mg tablets	60-mg tablets	30-mg tablets	10-mg tablets
99.17	54.21	28.51	19.06
94.31	55.62	26.25	8.83
95.92	57.40	25.92	9.08
94.55	57.51	28.62	
93.83	52.59	24.93	

(a) For each dosage, calculate the mean and standard deviation for the milligrams of morphine hydrochloride per tablet. (b) Assuming that $\bar{X}$ and s^2 are good approximations for μ and σ^2, and that the population is normally distributed, what percentage of tablets at each dosage level are expected to contain more than the nominal amount of morphine hydrochloride per tablet?

4. Daskalakis and co-workers recently evaluated several procedures for digesting the tissues of oysters and mussels prior to analyzing the samples for silver.[20] One of the methods used to evaluate the procedure is a spike recovery in which a known amount of silver is added to the tissue sample and the percent of the added silver found on analysis is reported. Ideally, spike recoveries should fall within the range 100 ± 15%. The results for one method are

106% 108% 92% 99% 104% 101% 93% 93%

Assuming that the spike recoveries are normally distributed, 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 kelvins, P is the pressure in atmospheres, and V is the volume in liters. In a typical analysis the following data are obtained (with estimated uncertainties in parentheses)

$g = 0.118\ (\pm 0.002)$
$R = 0.082056\ (\pm 0.000001)$
$T = 298.2\ (\pm 0.1)$
$P = 0.724\ (\pm 0.005)$
$V = 0.250\ (\pm 0.005)$

(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. A standard solution of Mn^{2+} was prepared by dissolving 0.250 g of Mn in 10 mL of concentrated HNO_3 (measured with a graduated cylinder). The resulting solution was quantitatively transferred to a 100-mL volumetric flask and diluted to volume with distilled water. A 10-mL aliquot of the solution was pipeted into a 500-mL volumetric flask and diluted to volume. (a) Express the concentration of Mn in parts per million, and estimate uncertainty by a propagation of uncertainty calculation. (b) Would the uncertainty in the solution's concentration be improved by using a pipet to measure the HNO_3, instead of a graduated cylinder?

7. Hydroscopic materials often are measured by the technique of weighing by difference. In this technique the material is placed in a sealed container and weighed. A portion of the material is removed, and the container and the remaining material are reweighed. The difference between the two masses gives the amount of material that was sampled. A solution of a hydroscopic material with a gram formula weight of 121.34 (±0.01) was prepared in the following manner. A sample of the compound and its container has a mass of 23.5811 g. A portion of the compound 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 g. Calculate the molarity of the solution, and estimate its uncertainty by a propagation of uncertainty calculation.

8. Show by a propagation of uncertainty calculation that the standard error of the mean for n determinations is given as $s/\sqrt{n}$.

9. What is the smallest mass that can be measured on an analytical balance with a tolerance of ±0.1 mg, such that the relative error is less than 0.1%?

10. 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?

11. A solution can be diluted 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, two, or three steps. Limiting yourself to glassware listed in Table 4.2, determine the proper combination of glassware to accomplish each dilution, and rank them in order of their most probable uncertainties.

12. Explain why changing all values in a data set by a constant amount will change $\bar{X}$ but will have no effect on s.

13. Obtain a sample of a metal from your instructor, and determine its density by one or both of the following methods:

Method A: Obtain the sample's mass with a balance. Calculate the sample's volume using appropriate linear dimensions.

Method B: Obtain 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 object, and reading the new volume. The difference in volumes is equal to the object's volume.

Determine the density at least five times. (a) Report the mean, the standard deviation, and the 95% confidence interval for your results. (b) Find the accepted value for the density of your metal, and determine the absolute and relative error for your experimentally determined density. (c) Use the propagation of uncertainty to determine the uncertainty for your chosen method. Are the results of this calculation consistent with your experimental results? If not, suggest some possible reasons for this disagreement.

14. How many carbon atoms must a molecule have if the mean number of ^{13}C atoms per molecule is 1.00? What percent of such molecules will have no atoms of ^{13}C?

15. In Example 4.10 we determined the probability that a molecule of cholesterol, $C_{27}H_{44}O$, had no atoms of ^{13}C. (a) Calculate the probability that a molecule of cholesterol, has one atom of ^{13}C. (b) What is the probability that a molecule of cholesterol will have two or more atoms of ^{13}C?

16. Berglund and Wichart investigated the quantitative determination of Cr in high-alloy steels by a potentiometric titration of Cr^{6+}.[21] Before titrating the steel was dissolved in acid and the chromium oxidized to Cr^{6+} by peroxydisulfate. Following are their results (%w/w Cr) for the analysis of a single 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?

17. Ketkar and co-workers developed a new analytical method for measuring trace levels of atmospheric gases.[22] The 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 $\alpha = 0.05$?

18. To test a spectrophotometer for its 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 a known 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 are analyzed with the following results

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 $\alpha = 0.01$.

19. Monna and co-workers studied the use of radioactive isotopes as a means of dating sediments collected from the bottom of lakes and estuaries.[23] 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. A 2.6540-g sample of an iron ore known to contain 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 method is used to determine the concentration of Fe in this solution, yielding results of 5840, 5770, 5650, and 5660 ppm. Determine whether there is a significant difference between the experimental mean and the expected value at $\alpha = 0.05$.

21. Horvat and colleagues investigated the application of atomic absorption spectroscopy to the analysis of Hg in coal fly ash.[24] Of particular interest was the development of an appropriate procedure for digesting the samples in order to release the Hg for analysis. As part of their study they tested several reagents for digesting samples. Results obtained with HNO_3 and with a 1 + 3 mixture of HNO_3 and HCl are shown here. All concentrations are given as nanograms of Hg per gram of sample.

HNO_3:	161	165	160	167	166	
1 + 3 HNO_3–HCl:	159	145	140	147	143	156

Determine whether there is a significant difference between these methods at $\alpha = 0.05$.

22. 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 chemistry in 1904 for the discovery of argon. An important turning point in the discovery of Ar was Rayleigh's experimental measurements of the density of N_2. Rayleigh approached this experiment in two ways: first by taking atmospheric air and removing any O_2 and H_2 that was present; and second, by chemically producing N_2 by decomposing nitrogen-containing compounds (NO, N_2O, and NH_4NO_3) and again removing any O_2 and H_2. His results for the density of N_2, published in *Proc. Roy. Soc.* **1894,** *LV,* 340 (publication 210), follow (all values are for grams of gas at equivalent volume, pressure, and temperature).

Atmospheric	2.31017	2.30986	2.31010	2.31001
Origin:	2.31024	2.31010	2.31028	
Chemical	2.30143	2.29890	2.29816	2.30182
Origin:	2.29869	2.29940	2.29849	2.29889

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

23. Gács and Ferraroli reported a new method for monitoring the concentration of SO_2 in air.[25] They compared their method with the standard method by sampling and analyzing urban air from a single location. Air samples were collected by drawing air through a collection solution for 6 min. Following is a summary of their results with SO_2 concentrations reported in microliters per cubic meter.

Standard method:	21.62	22.20	24.27	23.54	24.25	23.09	21.02
New method:	21.54	20.51	22.31	21.30	24.62	25.72	21.54

Using an appropriate statistical test, determine whether there is any significant difference between the standard and new methods at $\alpha = 0.05$.

24. The accuracy of a spectrophotometer can be checked by measuring absorbances for a series of standard dichromate solutions that can be obtained in sealed cuvettes from the National Institute of Standards and Technology. Absorbances are measured at 257 nm and compared with 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:	1	2	3	4	5
Measured absorbance:	0.2872	0.5773	0.8674	1.1623	1.4559
Accepted absorbance:	0.2871	0.5760	0.8677	1.1608	1.4565

25. Maskarinec and associates investigated the stability of volatile organics in environmental water samples.[26] 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. Results (in days) for surface waters follow.

	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-tetrachloroethane	33	85
tetrachloroethene	41	63
toluene	32	94
chlorobenzene	36	86

Determine whether there is a significant difference in the effectiveness of the two preservatives at $\alpha = 0.10$.

26. Using X-ray diffraction, Karstang and Kvalhein reported a new method for determining the weight percent of kalonite in complex clay minerals.[27] To test the method, nine samples containing known amounts of kalonite were prepared and analyzed. The results (as %w/w kalonite) are shown.

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 $\alpha = 0.05$.

27. Mizutani and colleagues reported the development of a new method for the analysis of l-malate.[28] As part of their study they analyzed a series of beverages using both their method and a standard spectrophotometric procedure based on a clinical kit purchased from Boerhinger Scientific. A summary follows of their results (in parts per million).

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 $\alpha = 0.05$.

28. Alexiev and associates describe an improved photometric method for the determination of Fe^{3+} based on its catalytic effect on the oxidation of sulphanilic acid by KIO_4.[29] As part of their study the concentration of Fe^{3+} in human serum samples was determined by the proposed method and the standard method. Following are the results, with concentrations in micromoles/L.

Sample	Proposed 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	13.43

Determine whether there is a significant difference between the two methods at $\alpha = 0.05$.

29. The following data were collected during a study of the concentration of Zn in samples drawn from several locations in Lake Erie (all concentrations in parts per million).

Location	$[Zn^{2+}]$ at the air–water interface	$[Zn^{2+}]$ at the sediment–water interface
1	0.430	0.415
2	0.266	0.238
3	0.567	0.390
4	0.531	0.410
5	0.707	0.605
6	0.716	0.609

Determine whether there is a significant difference between the concentration of Zn^{2+} at the air–water interface and the sediment–water interface at $\alpha = 0.05$.

30. Ten laboratories were asked to determine the concentration of an analyte A in three standard test samples. Following are the results, in parts per million.[30]

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	16.3
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 $\alpha = 0.05$.

31. When copper metal and powdered sulfur are placed in a crucible and ignited, the product is a sulfide with an empirical formula of Cu_xS. The value of x can be determined by weighing the Cu and S before ignition, and finding the mass of Cu_xS when the reaction is complete. Following are the Cu/S ratios from 62 such experiments.

1.764	1.838	1.865	1.866	1.872	1.877	1.890	1.891	1.891
1.897	1.899	1.900	1.906	1.908	1.910	1.911	1.916	1.919
1.920	1.922	1.927	1.931	1.935	1.936	1.936	1.937	1.939
1.939	1.940	1.941	1.941	1.942	1.943	1.948	1.953	1.955
1.957	1.957	1.957	1.959	1.962	1.963	1.963	1.963	1.966
1.968	1.969	1.973	1.975	1.976	1.977	1.981	1.981	1.988
1.993	1.993	1.995	1.995	1.995	2.017	2.029	2.042	

(a) Calculate the mean and standard deviation for these data. (b) Construct a histogram for this data set. From a visual inspection of your histogram, do 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 $\bar{X} \pm 1s$? Does this support your answer to the previous

question? (d) Assuming that $\bar{X}$ and σ^2 are good approximations for μ and σ^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 are normally distributed? (e) It has been reported that this method for preparing copper sulfide results in a nonstoichiometric compound with a Cu/S ratio of less than 2. Determine if the mean value for these data is significantly less than 2 at a significance level of $\alpha = 0.01$.

4L SUGGESTED READINGS

A more comprehensive discussion of the analysis of data, covering all topics considered in this chapter as well as additional material, can be found in any textbook on statistics or data analysis; following are several such texts.

Anderson, R. L. *Practical Statistics for Analytical Chemists.* Van Nostrand Reinhold: New York, 1987.

Graham, R. C. *Data Analysis for the Chemical Sciences.* VCH Publishers: New York, 1993.

Mark, H.; Workman, J. *Statistics in Spectroscopy.* Academic Press: Boston, 1991.

Mason, R. L.; Gunst, R. F.; Hess, J. L. *Statistical Design and Analysis of Experiments.* Wiley: New York, 1989.

Miller, J. C.; Miller, J. N. *Statistics for Analytical Chemistry,* 3rd ed. Ellis Horwood PTR Prentice-Hall: New York, 1993.

Sharaf, M. H.; Illman, D. L.; Kowalski, B. R. *Chemometrics.* Wiley-Interscience: New York, 1986.

The difference between precision and accuracy and a discussion of indeterminate and determinate sources of error is covered in the following paper.

Treptow, R. S. "Precision and Accuracy in Measurements," *J. Chem. Educ.* **1998,** *75,* 992–995.

The detection of outliers, particularly when working with a small number of samples, is discussed in the following papers.

Efstathiou, C. "Stochastic Calculation of Critical *Q*-Test Values for the Detection of Outliers in Measurements," *J. Chem. Educ.* **1992,** *69,* 773–736.

Kelly, P. C. "Outlier Detection in Collaborative Studies," *Anal. Chem.* **1990,** *73,* 58–64.

Mitschele, J. "Small Sample Statistics," *J. Chem. Educ.* **1991,** *68,* 470–473.

The following papers provide additional information on error and uncertainty, including the propagation of uncertainty.

Andraos, J. "On the Propagation of Statistical Errors for a Function of Several Variables," *J. Chem. Educ.* **1996,** *73,* 150–154.

Donato, H.; Metz, C. "A Direct Method for the Propagation of Error Using a Personal Computer Spreadsheet Program," *J. Chem. Educ.* **1988,** *65,* 867–868.

Gordon, R.; Pickering, M.; Bisson, D. "Uncertainty Analysis by the 'Worst Case' Method," *J. Chem. Educ.* **1984,** *61,* 780–781.

Guare, C. J. "Error, Precision and Uncertainty," *J. Chem. Educ.* **1991,** *68,* 649–652.

Guedens, W. J.; Yperman, J.; Mullens, J.; et al. "Statistical Analysis of Errors: A Practical Approach for an Undergraduate Chemistry Lab," Part 1. The Concept, *J. Chem. Educ.* **1993,** *70,* 776–779; Part 2. Some Worked Examples, *J. Chem. Educ.* **1993,** *70,* 838–841.

Heydorn, K. "Detecting Errors in Micro and Trace Analysis by Using Statistics," *Anal. Chim. Acta* **1993,** *283,* 494–499.

Taylor, B. N.; Kuyatt, C. E. "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results," *NIST Technical Note* 1297, **1994.**

A further discussion of detection limits is found in

Currie, L. A., ed. *Detection in Analytical Chemistry: Importance, Theory and Practice.* American Chemical Society: Washington, DC, 1988.

4M REFERENCES

1. Goedhart, M. J.; Verdonk, A. H. *J. Chem. Educ.* **1991,** *68,* 1005–1009.
2. Rousseeuw, P. J. *J. Chemom.* **1991,** *5,* 1–20.
3. Ellison, S.; Wegscheider, W.; Williams, A. *Anal. Chem.* **1997,** *69,* 607A–613A.
4. Shoemaker, D. P.; Garland, C. W.; Nibler, J. W. *Experiments in Physical Chemistry,* 5th ed. McGraw-Hill: New York, **1989,** pp. 55–63.
5. Lam, R. B.; Isenhour, T. L. *Anal. Chem.* **1980,** *52,* 1158–1161.
6. Mark, H.; Workman, J. *Spectroscopy,* **1988,** *3(1),* 44–48.
7. Winn, R. L. *Statistics for Scientists and Engineers,* Prentice-Hall: Englewood Cliffs, NJ, 1964; pp. 165–174.
8. Mark, H.; Workman, J. *Spectroscopy,* **1989,** *4(3),* 56–58.
9. Marecek, V.; Janchenova, H.; Brezina, M.; et al. *Anal. Chim. Acta* **1991,** *244,* 15–19.
10. Rorabacher, D. B. *Anal. Chem.* **1991,** *63,* 139–146.
11. Deming, W. E. *Statistical Adjustment of Data.* Wiley: New York, 1943 (republished by Dover: New York, 1961); p. 171.
12. Richardson, T. H. *J. Chem. Educ.* **1991,** *68,* 310–311.
13. 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, DC, 1988.
14. Long, G. L.; Winefordner, J. D. *Anal. Chem.* **1983,** *55,* 712A–724A.

15. "Guidelines for Data Acquisition and Data Quality Control Evaluation in Environmental Chemistry," *Anal. Chem.* **1980,** *52,* 2242–2249.
16. (a) Ferrus, R.; Egea, M. R. *Anal. Chim. Acta* **1994,** *287,* 119–145; (b) Glaser, J. A.; Foerst, D. L.; McKee, G. D.; et al. *Environ. Sci. Technol.* **1981,** *15,* 1426–1435; (c) Boumans, P. W. J. M. *Anal. Chem.* **1994,** *66,* 459A–467A; (d) Kimbrough, D. E.; Wakakuwa, J. *Environ. Sci. Technol.* **1994,** *28,* 338–345; (e) Currie, L. A. *Anal. Chem.* **1968,** *40,* 586–593.
17. (a) Rogers, L. B. *J. Chem. Educ.* **1986,** *63,* 3–6; (b) Mark, H.; Workman, J. *Spectroscopy,* **1989,** *4(1),* 52–55.
18. Simonian, M. H.; Dinh, S.; Fray, L. A. *Spectroscopy* **1993,** *8(6),* 37–47.
19. Salem, I. I.; Galan, A. C. *Anal. Chim. Acta* **1993,** *283,* 334–337.
20. Daskalakis, K. D.; O'Connor, T. P.; Crecelius, E. A. *Environ. Sci. Technol.* **1997,** *31,* 2303–2306.
21. Berglund, B.; Wichardt, C. *Anal. Chim. Acta* **1990,** *236,* 399–410.
22. Ketkar, S. N.; Dulak, J. G.; Dheandhanou, S.; et al. *Anal. Chim. Acta.* **1991,** *245,* 267–270.
23. Monna, F.; Mathieu, D.; Marques Jr., et al. *Anal. Chim. Acta* **1996,** *330,* 107–116.
24. Horvat, M.; Lupsina, V.; Pihlar, B. *Anal. Chim. Acta.* **1991,** *243,* 71–79.
25. Gács, I.; Ferraroli, R. *Anal. Chim. Acta* **1992,** *269,* 177–185.
26. Maskarinec, M. P.; Johnson, L. H.; Holladay, S. K.; et al. *Environ. Sci. Technol.* **1990,** *24,* 1665–1670.
27. Karstang, T. V.; Kvalhein, O. M. *Anal. Chem.* **1991,** *63,* 767–772.
28. Mizutani, F.; Yabuki, S.; Asai, M. *Anal. Chim. Acta.* **1991,** *245,* 145–150.
29. Alexiev, A.; Rubino, S.; Deyanova, M.; et al. *Anal. Chem. Acta* **1994,** *295,* 211–219.
30. These data are adapted from Steiner, E. H. "Planning and Analysis of Results of Collaborative Tests" published in *Statistical Manual of the Association of Official Analytical Chemists,* Association of Official Analytical Chemists: Washington, DC, 1975.

Chapter 5

Calibrations, Standardizations, and Blank Corrections

In Chapter 3 we introduced a relationship between the measured signal, S_{meas}, and the absolute amount of analyte

$$S_{meas} = kn_A + S_{reag} \quad \textbf{5.1}$$

or the relative amount of analyte in a sample

$$S_{meas} = kC_A + S_{reag} \quad \textbf{5.2}$$

where n_A is the moles of analyte, C_A is the analyte's concentration, k is the method's sensitivity, and S_{reag} is the contribution to S_{meas} from constant errors introduced by the reagents used in the analysis. To obtain an accurate value for n_A or C_A it is necessary to avoid determinate errors affecting S_{meas}, k, and S_{reag}. This is accomplished by a combination of calibrations, standardizations, and reagent blanks.

5A Calibrating Signals

Signals are measured using equipment or instruments that must be properly calibrated if S_{meas} is to be free of determinate errors. Calibration is accomplished against a standard, adjusting S_{meas} until it agrees with the standard's known signal. Several common examples of calibration are discussed here.

When the signal is a measurement of mass, S_{meas} is determined with an analytical balance. Before a balance can be used, it must be calibrated against a reference weight meeting standards established by either the National Institute for Standards and Technology or the American Society for Testing and Materials. With an electronic balance the sample's mass is determined by the current required to generate an upward electromagnetic force counteracting the sample's downward gravitational force. The balance's calibration procedure invokes an internally programmed calibration routine specifying the reference weight to be used. The reference weight is placed on the balance's weighing pan, and the relationship between the displacement of the weighing pan and the counteracting current is automatically adjusted.

Calibrating a balance, however, does not eliminate all sources of determinate error. Due to the buoyancy of air, an object's weight in air is always lighter than its weight in vacuum. If there is a difference between the density of the object being weighed and the density of the weights used to calibrate the balance, then a correction to the object's weight must be made.[1] An object's true weight in vacuo, W_v, is related to its weight in air, W_a, by the equation

$$W_v = W_a \times \left[1 + \left(\frac{1}{D_o} - \frac{1}{D_w}\right) \times 0.0012\right]$$

where D_o is the object's density, D_w is the density of the calibration weight, and 0.0012 is the density of air under normal laboratory conditions (all densities are in units of g/cm^3). Clearly the greater the difference between D_o and D_w the more serious the error in the object's measured weight.

The buoyancy correction for a solid is small, and frequently ignored. It may be significant, however, for liquids and gases of low density. This is particularly important when calibrating glassware. For example, a volumetric pipet is calibrated by carefully filling the pipet with water to its calibration mark, dispensing the water into a tared beaker and determining the mass of water transferred. After correcting for the buoyancy of air, the density of water is used to calculate the volume of water dispensed by the pipet.

Example

EXAMPLE 5.1

A 10-mL volumetric pipet was calibrated following the procedure just outlined, using a balance calibrated with brass weights having a density of 8.40 g/cm^3. At 25 °C the pipet was found to dispense 9.9736 g of water. What is the actual volume dispensed by the pipet?

SOLUTION

At 25 °C the density of water is 0.99705 g/cm^3. The water's true weight, therefore, is

$$W_v = 9.9736 \text{ g} \times \left[1 + \left(\frac{1}{0.99705} - \frac{1}{8.40}\right) \times 0.0012\right] = 9.9842 \text{ g}$$

and the actual volume of water dispensed by the pipet is

$$\frac{9.9842\ \text{g}}{0.99705\ \text{g/cm}^3} = 10.014\ \text{cm}^3 = 10.014\ \text{mL}$$

If the buoyancy correction is ignored, the pipet's volume is reported as

$$\frac{9.9736\ \text{g}}{0.99705\ \text{g/cm}^3} = 10.003\ \text{cm}^3 = 10.003\ \text{mL}$$

introducing a negative determinate error of –0.11%.

Balances and volumetric glassware are examples of laboratory equipment. Laboratory instrumentation also must be calibrated using a standard providing a known response. For example, a spectrophotometer's accuracy can be evaluated by measuring the absorbance of a carefully prepared solution of 60.06 ppm $K_2Cr_2O_7$ in 0.0050 M H_2SO_4, using 0.0050 M H_2SO_4 as a reagent blank.[2] The spectrophotometer is considered calibrated if the resulting absorbance at a wavelength of 350.0 nm is 0.640 ± 0.010 absorbance units. Be sure to read and carefully follow the calibration instructions provided with any instrument you use.

5B Standardizing Methods

The American Chemical Society's Committee on Environmental Improvement defines standardization as the process of determining the relationship between the measured signal and the amount of analyte.[3] A method is considered standardized when the value of k in equation 5.1 or 5.2 is known.

In principle, it should be possible to derive the value of k for any method by considering the chemical and physical processes responsible for the signal. Unfortunately, such calculations are often of limited utility due either to an insufficiently developed theoretical model of the physical processes or to nonideal chemical behavior. In such situations the value of k must be determined experimentally by analyzing one or more standard solutions containing known amounts of analyte. In this section we consider several approaches for determining the value of k. For simplicity we will assume that S_{reag} has been accounted for by a proper reagent blank, allowing us to replace S_{meas} in equations 5.1 and 5.2 with the signal for the species being measured.

5B.1 Reagents Used as Standards

The accuracy of a standardization depends on the quality of the reagents and glassware used to prepare standards. For example, in an acid–base titration, the amount of analyte is related to the absolute amount of titrant used in the analysis by the stoichiometry of the chemical reaction between the analyte and the titrant. The amount of titrant used is the product of the signal (which is the volume of titrant) and the titrant's concentration. Thus, the accuracy of a titrimetric analysis can be no better than the accuracy to which the titrant's concentration is known.

primary reagent
A reagent of known purity that can be used to make a solution of known concentration.

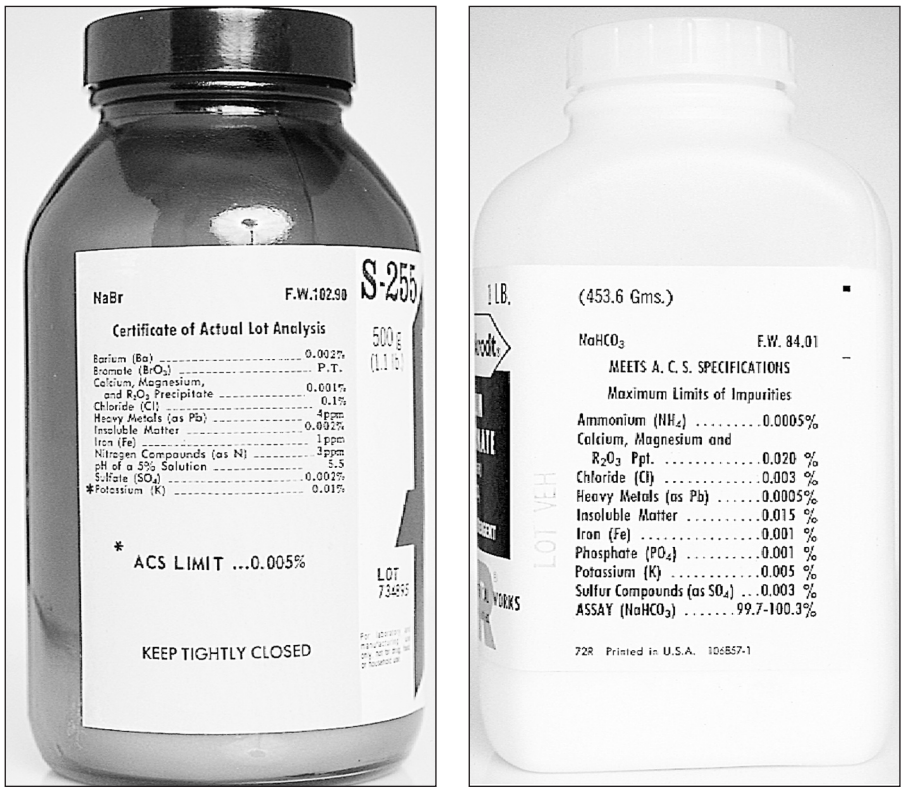
Primary Reagents Reagents used as standards are divided into primary reagents and secondary reagents. A **primary reagent** can be used to prepare a standard containing an accurately known amount of analyte. For example, an accurately weighed sample of 0.1250 g $K_2Cr_2O_7$ contains exactly 4.249×10^{-4} mol of $K_2Cr_2O_7$. If this

same sample is placed in a 250-mL volumetric flask and diluted to volume, the concentration of the resulting solution is exactly 1.700×10^{-3} M. A primary reagent must have a known stoichiometry, a known purity (or assay), and be stable during long-term storage both in solid and solution form. Because of the difficulty in establishing the degree of hydration, even after drying, hydrated materials usually are not considered primary reagents. Reagents not meeting these criteria are called **secondary reagents.** The purity of a secondary reagent in solid form or the concentration of a standard prepared from a secondary reagent must be determined relative to a primary reagent. Lists of acceptable primary reagents are available.[4] Appendix 2 contains a selected listing of primary standards.

secondary reagent
A reagent whose purity must be established relative to a primary reagent.

Other Reagents Preparing a standard often requires additional substances that are not primary or secondary reagents. When a standard is prepared in solution, for example, a suitable solvent and solution matrix must be used. Each of these solvents and reagents is a potential source of additional analyte that, if unaccounted for, leads to a determinate error. If available, **reagent grade** chemicals conforming to standards set by the American Chemical Society should be used.[5] The packaging label included with a reagent grade chemical (Figure 5.1) lists either the maximum allowed limit for specific impurities or provides the actual assayed values for the impurities as reported by the manufacturer. The purity of a reagent grade chemical can be improved by purification or by conducting a more accurate assay. As discussed later in the chapter, contributions to S_{meas} from impurities in the sample matrix can be compensated for by including an appropriate blank determination in the analytical procedure.

reagent grade
Reagents conforming to standards set by the American Chemical Society.



(a) (b)

Figure 5.1
Examples of typical packaging labels from reagent grade chemicals. Label (a) provides the actual lot assay for the reagent as determined by the manufacturer. Note that potassium has been flagged with an asterisk (*) because its assay exceeds the maximum limit established by the American Chemical Society (ACS). Label (b) does not provide assayed values, but indicates that the reagent meets the specifications of the ACS for the listed impurities. An assay for the reagent also is provided.

Preparing Standard Solutions Solutions of primary standards generally are prepared in class A volumetric glassware to minimize determinate errors. Even so, the relative error in preparing a primary standard is typically ±0.1%. The relative error can be improved if the glassware is first calibrated as described in Example 5.1. It also is possible to prepare standards gravimetrically by taking a known mass of standard, dissolving it in a solvent, and weighing the resulting solution. Relative errors of ±0.01% can typically be achieved in this fashion.

It is often necessary to prepare a series of standard solutions, each with a different concentration of analyte. Such solutions may be prepared in two ways. If the range of concentrations is limited to only one or two orders of magnitude, the solutions are best prepared by transferring a known mass or volume of the pure standard to a volumetric flask and diluting to volume. When working with larger concentration ranges, particularly those extending over more than three orders of magnitude, standards are best prepared by a serial dilution from a single stock solution. In a serial dilution a volume of a concentrated stock solution, which is the first standard, is diluted to prepare a second standard. A portion of the second standard is then diluted to prepare a third standard, and the process is repeated until all necessary standards have been prepared. Serial dilutions must be prepared with extra care because a determinate error in the preparation of any single standard is passed on to all succeeding standards.

5B.2 Single-Point versus Multiple-Point Standardizations*

single-point standardization
Any standardization using a single standard containing a known amount of analyte.

The simplest way to determine the value of k in equation 5.2 is by a **single-point standardization.** A single standard containing a known concentration of analyte, C_S, is prepared and its signal, S_{stand}, is measured. The value of k is calculated as

$$k = \frac{S_{stand}}{C_S} \qquad \textbf{5.3}$$

A single-point standardization is the least desirable way to standardize a method. When using a single standard, all experimental errors, both determinate and indeterminate, are carried over into the calculated value for k. Any uncertainty in the value of k increases the uncertainty in the analyte's concentration. In addition, equation 5.3 establishes the standardization relationship for only a single concentration of analyte. Extending equation 5.3 to samples containing concentrations of analyte different from that in the standard assumes that the value of k is constant, an assumption that is often not true.[6] Figure 5.2 shows how assuming a constant value of k may lead to a determinate error. Despite these limitations, single-point standardizations are routinely used in many laboratories when the analyte's range of expected concentrations is limited. Under these conditions it is often safe to assume that k is constant (although this assumption should be verified experimentally). This is the case, for example, in clinical laboratories where many automated analyzers use only a single standard.

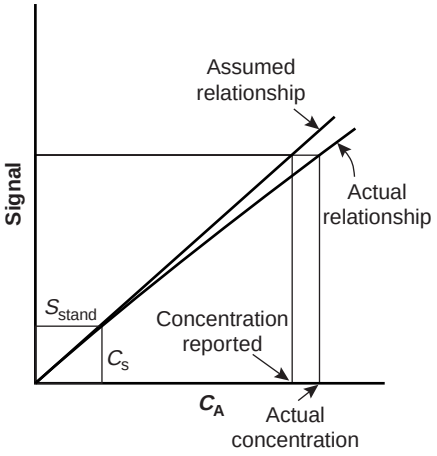


Figure 5.2
Example showing how an improper use of a single-point standardization can lead to a determinate error in the reported concentration of analyte.

The preferred approach to standardizing a method is to prepare a series of standards, each containing the analyte at a different concentration. Standards are chosen such that they bracket the expected range for the

*The following discussion of standardizations assumes that the amount of analyte is expressed as a concentration. It also applies, however, when the absolute amount of analyte is given in grams or moles.

analyte's concentration. Thus, a **multiple-point standardization** should use at least three standards, although more are preferable. A plot of S_{stand} versus C_S is known as a calibration curve. The exact standardization, or calibration relationship, is determined by an appropriate curve-fitting algorithm.* Several approaches to standardization are discussed in the following sections.

multiple-point standardization
Any standardization using two or more standards containing known amounts of analyte.

5B.3 External Standards

The most commonly employed standardization method uses one or more **external standards** containing known concentrations of analyte. These standards are identified as external standards because they are prepared and analyzed separately from the samples.

external standard
A standard solution containing a known amount of analyte, prepared separately from samples containing the analyte.

A quantitative determination using a single external standard was described at the beginning of this section, with k given by equation 5.3. Once standardized, the concentration of analyte, C_A, is given as

$$C_A = \frac{S_{samp}}{k} \qquad \textbf{5.4}$$

Example

EXAMPLE 5.2

A spectrophotometric method for the quantitative determination of Pb^{2+} levels in blood yields an S_{stand} of 0.474 for a standard whose concentration of lead is 1.75 ppb. How many parts per billion of Pb^{2+} occur in a sample of blood if S_{samp} is 0.361?

SOLUTION

Equation 5.3 allows us to calculate the value of k for this method using the data for the standard

$$k = \frac{S_{stand}}{C_S} = \frac{0.474}{1.75 \text{ ppb}} = 0.2709 \text{ ppb}^{-1}$$

Once k is known, the concentration of Pb^{2+} in the sample of blood can be calculated using equation 5.4

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

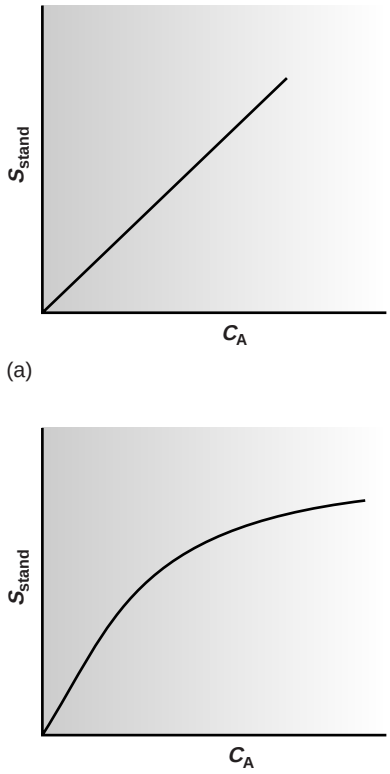


Figure 5.3
Examples of (a) straight-line and (b) curved normal calibration curves.

A multiple-point external standardization is accomplished by constructing a calibration curve, two examples of which are shown in Figure 5.3. Since this is the most frequently employed method of standardization, the resulting relationship often is called a **normal calibration curve.** When the calibration curve is a linear (Figure 5.3a), the slope of the line gives the value of k. This is the most desirable situation since the method's sensitivity remains constant throughout the standard's concentration range. When the calibration curve is nonlinear, the method's sensitivity is a function of the analyte's concentration. In Figure 5.3b, for example, the value of k is greatest when the analyte's concentration is small and decreases continuously as the amount of analyte is increased. The value of k at any point along the calibration curve is given by the slope at that point. In

normal calibration curve
A calibration curve prepared using several external standards.

*Linear regression, also known as the method of least squares, is covered in Section 5C.

either case, the calibration curve provides a means for relating S_{samp} to the analyte's concentration.

Colorplate 1 shows an example of a set of external standards and their corresponding normal calibration curve.

Example

EXAMPLE 5.3

A second spectrophotometric method for the quantitative determination of Pb^{2+} levels in blood gives a linear normal calibration curve for which

$$S_{stand} = (0.296\ ppb^{-1}) \times C_S + 0.003$$

What is the Pb^{2+} level (in ppb) in a sample of blood if S_{samp} is 0.397?

SOLUTION

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

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

It is worth noting that the calibration equation in this problem includes an extra term that is not in equation 5.3. Ideally, we expect the calibration curve to give a signal of zero when C_S is zero. This is the purpose of using a reagent blank to correct the measured signal. The extra term of +0.003 in our calibration equation results from uncertainty in measuring the signal for the reagent blank and the standards.

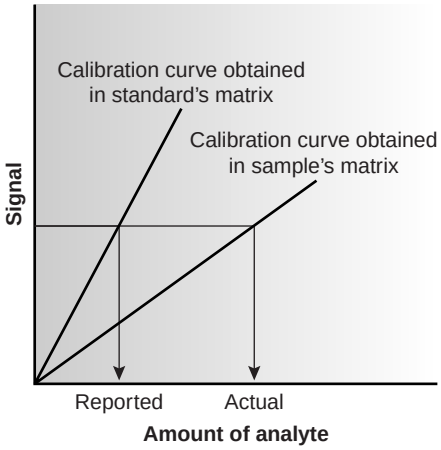


Figure 5.4
Effect of the sample's matrix on a normal calibration curve.

An external standardization allows a related series of samples to be analyzed using a single calibration curve. This is an important advantage in laboratories where many samples are to be analyzed or when the need for a rapid throughput of samples is critical. Not surprisingly, many of the most commonly encountered quantitative analytical methods are based on an external standardization.

There is a serious limitation, however, to an external standardization. The relationship between S_{stand} and C_S in equation 5.3 is determined when the analyte is present in the external standard's matrix. In using an external standardization, we assume that any difference between the matrix of the standards and the sample's matrix has no effect on the value of k. A proportional determinate error is introduced when differences between the two matrices cannot be ignored. This is shown in Figure 5.4, where the relationship between the signal and the amount of analyte is shown for both the sample's matrix and the standard's matrix. In this example, using a normal calibration curve results in a negative determinate error. When matrix problems are expected, an effort is made to match the matrix of the standards to that of the sample. This is known as **matrix matching.** When the sample's matrix is unknown, the matrix effect must be shown to be negligible, or an alternative method of standardization must be used. Both approaches are discussed in the following sections.

matrix matching
Adjusting the matrix of an external standard so that it is the same as the matrix of the samples to be analyzed.

method of standard additions
A standardization in which aliquots of a standard solution are added to the sample.

5B.4 Standard Additions

The complication of matching the matrix of the standards to that of the sample can be avoided by conducting the standardization in the sample. This is known as the **method of standard additions.** The simplest version of a standard addi-

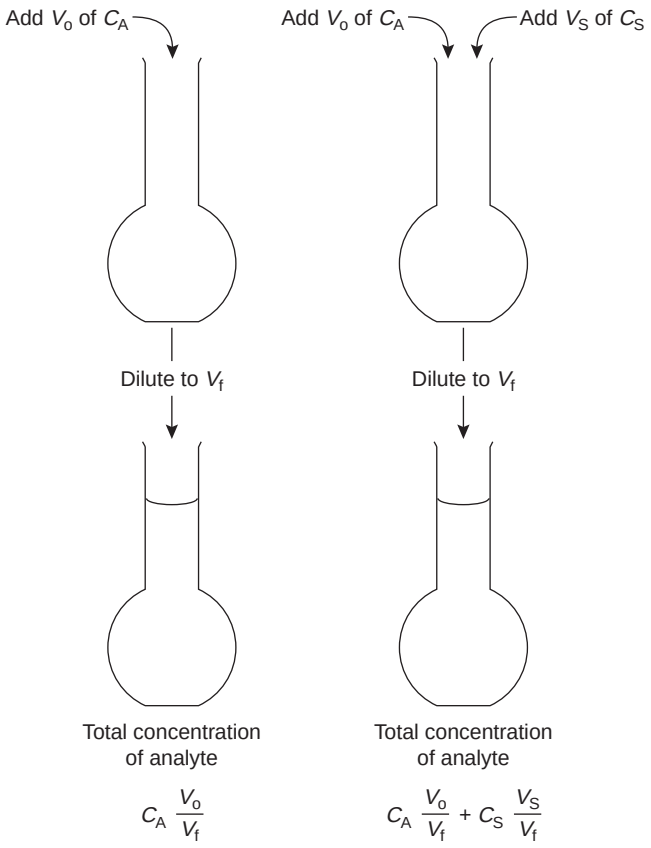


Figure 5.5
Illustration showing the method of standard additions in which separate aliquots of sample are diluted to the same final volume. One aliquot of sample is spiked with a known volume of a standard solution of analyte before diluting to the final volume.

tion is shown in Figure 5.5. A volume, V_o, of sample is diluted to a final volume, V_f, and the signal, S_{samp} is measured. A second identical **aliquot** of sample is spiked with a volume, V_s, of a standard solution for which the analyte's concentration, C_S, is known. The spiked sample is diluted to the same final volume and its signal, S_{spike}, is recorded. The following two equations relate S_{samp} and S_{spike} to the concentration of analyte, C_A, in the original sample

aliquot
A portion of a solution.

$$S_{samp} = kC_A \frac{V_o}{V_f} \quad \textbf{5.5}$$

$$S_{spike} = k\left(C_A \frac{V_o}{V_f} + C_S \frac{V_s}{V_f}\right) \quad \textbf{5.6}$$

where the ratios V_o/V_f and V_s/V_f account for the dilution. As long as V_s is small relative to V_o, the effect of adding the standard to the sample's matrix is insignificant, and the matrices of the sample and the spiked sample may be considered identical. Under these conditions the value of k is the same in equations 5.5 and 5.6. Solving both equations for k and equating gives

$$\frac{S_{samp}}{C_A(V_o/V_f)} = \frac{S_{spike}}{C_A(V_o/V_f) + C_S(V_s/V_f)} \quad \textbf{5.7}$$

Equation 5.7 can be solved for the concentration of analyte in the original sample.

Example

EXAMPLE 5.4

A third spectrophotometric method for the quantitative determination of the concentration of Pb^{2+} in blood yields an S_{samp} of 0.193 for a 1.00-mL sample of blood that has been diluted to 5.00 mL. A second 1.00-mL sample is spiked with 1.00 µL of a 1560-ppb Pb^{2+} standard and diluted to 5.00 mL, yielding an S_{spike} of 0.419. Determine the concentration of Pb^{2+} in the original sample of blood.

SOLUTION

The concentration of Pb^{2+} in the original sample of blood can be determined by making appropriate substitutions into equation 5.7 and solving for C_A. Note that all volumes must be in the same units, thus V_s is converted from 1.00 µL to 1.00×10^{-3} mL.

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

$$\frac{0.193}{0.200C_A} = \frac{0.419}{0.200C_A + 0.312\text{ ppb}}$$

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

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

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

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

It also is possible to make a standard addition directly to the sample after measuring S_{samp} (Figure 5.6). In this case, the final volume after the standard addition is $V_o + V_s$ and equations 5.5–5.7 become

$$S_{samp} = kC_A$$

$$S_{spike} = k\left(C_A\frac{V_o}{V_o + V_s} + C_S\frac{V_s}{V_o + V_s}\right) \qquad \textbf{5.8}$$

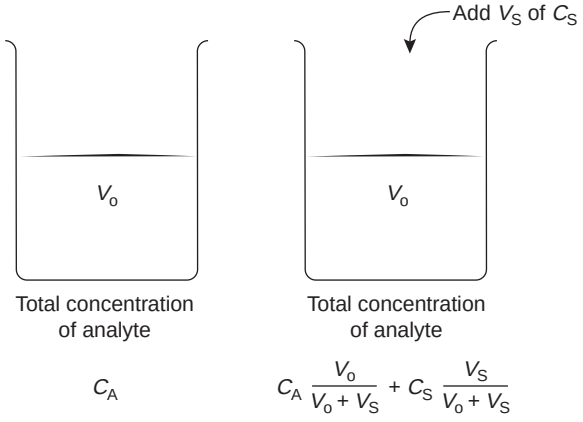


Figure 5.6
Illustration showing an alternative form of the method of standard additions. In this case a sample containing the analyte is spiked with a known volume of a standard solution of analyte without further diluting either the sample or the spiked sample.

$$\frac{S_{samp}}{C_A} = \frac{S_{spike}}{C_A[V_o/(V_o + V_s)] + C_S[V_s/(V_o + V_s)]} \qquad \textbf{5.9}$$

Example

EXAMPLE 5.5

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

SOLUTION

The concentration of Pb^{2+} in the original sample of blood can be determined by making appropriate substitutions into equation 5.9 and solving for C_A.

$$\frac{0.712}{C_A} = \frac{1.546}{C_A\left[\frac{5.00\text{ mL}}{(5.00\text{ mL} + 5.00 \times 10^{-3}\text{ mL})}\right] + 1560\text{ ppb}\left[\frac{5.00 \times 10^{-3}\text{ mL}}{(5.00\text{ mL} + 5.00 \times 10^{-3}\text{ mL})}\right]}$$

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

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

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

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

The single-point standard additions outlined in Examples 5.4 and 5.5 are easily adapted to a multiple-point standard addition by preparing a series of spiked samples containing increasing amounts of the standard. A calibration curve is prepared by plotting S_{spike} versus an appropriate measure of the amount of added standard. Figure 5.7 shows two examples of a standard addition calibration curve based on equation 5.6. In Figure 5.7(a) S_{spike} is plotted versus the volume of the standard solution spikes, V_s. When k is constant, the calibration curve is linear, and it is easy to show that the x-intercept's absolute value is $C_A V_o/C_S$.

Colorplate 2 shows an example of a set of standard additions and their corresponding standard additions calibration curve.

Example

EXAMPLE 5.6

Starting with equation 5.6, show that the equations for the slope, y-intercept, and x-intercept in Figure 5.7(a) are correct.

SOLUTION

We begin by rewriting equation 5.6 as

$$S_{spike} = \frac{kC_A V_o}{V_f} + \frac{kC_S}{V_f} \times V_s$$

which is in the form of the linear equation

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

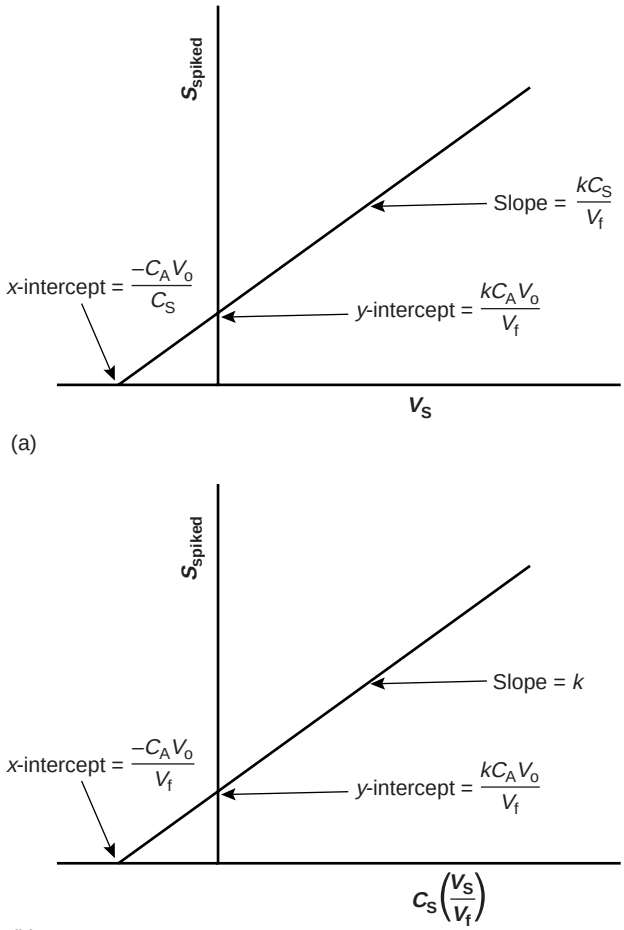


Figure 5.7
Examples of calibration curves for the method of standard additions. In (a) the signal is plotted versus the volume of the added standard, and in (b) the signal is plotted versus the concentration of the added standard after dilution.

where Y is S_{spike} and X is V_s. The slope of the line, therefore, is kC_S/V_f, and the y-intercept is kC_AV_o/V_f. The x-intercept is the value of X when Y is 0, or

$$0 = \frac{kC_AV_o}{V_f} + \frac{kC_S}{V_f} \times (x\text{-intercept})$$

$$x\text{-intercept} = -\frac{(kC_AV_o/V_f)}{(kC_S/V_f)} = -\frac{C_AV_o}{C_S}$$

Thus, the absolute value of the x-intercept is C_AV_o/C_S.

Since both V_o and C_S are known, the x-intercept can be used to calculate the analyte's concentration.

Example

EXAMPLE 5.7

A fifth spectrophotometric method for the quantitative determination of the concentration of Pb^{2+} in blood uses a multiple-point standard addition based on equation 5.6. The original blood sample has a volume of 1.00 mL, and the standard used for spiking the sample has a concentration of 1560 ppb Pb^{2+}. All samples were diluted to 5.00 mL before measuring the signal. A calibration curve of S_{spike} versus V_s is described by

$$S_{spike} = 0.266 + 312\ mL^{-1} \times V_s$$

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

SOLUTION

To find the x-intercept we let S_{spike} equal 0

$$0 = 0.266 + 312\ mL^{-1} \times (x\text{-intercept})$$

and solve for the x-intercept's absolute value, giving a value of 8.526×10^{-4} mL. Thus

$$x\text{-intercept} = 8.526 \times 10^{-4}\ mL = \frac{C_A V_o}{C_S} = \frac{C_A \times (1.00\ mL)}{1560\ ppb}$$

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

Figure 5.7(b) shows the relevant relationships when S_{spike} is plotted versus the concentrations of the spiked standards after dilution. Standard addition calibration curves based on equation 5.8 are also possible.

Since a standard additions calibration curve is constructed in the sample, it cannot be extended to the analysis of another sample. Each sample, therefore, requires its own standard additions calibration curve. This is a serious drawback to the routine application of the method of standard additions, particularly in laboratories that must handle many samples or that require a quick turnaround time. For example, suppose you need to analyze ten samples using a three-point calibration curve. For a normal calibration curve using external standards, only 13 solutions need to be analyzed (3 standards and 10 samples). Using the method of standard additions, however, requires the analysis of 30 solutions, since each of the 10 samples must be analyzed three times (once before spiking and two times after adding successive spikes).

The method of standard additions can be used to check the validity of an external standardization when matrix matching is not feasible. To do this, a normal calibration curve of S_{stand} versus C_S is constructed, and the value of k is determined from its slope. A standard additions calibration curve is then constructed using equation 5.6, plotting the data as shown in Figure 5.7(b). The slope of this standard additions calibration curve gives an independent determination of k. If the two values of k are identical, then any difference between the sample's matrix and that of the external standards can be ignored. When the values of k are different, a proportional determinate error is introduced if the normal calibration curve is used.

5B.5 Internal Standards

The successful application of an external standardization or the method of standard additions, depends on the analyst's ability to handle samples and standards reproducibly. When a procedure cannot be controlled to the extent that all samples and standards are treated equally, the accuracy and precision of the standardization may suffer. For example, if an analyte is present in a volatile solvent, its concentration will increase if some solvent is lost to evaporation. Suppose that you have a sample and a standard with identical concentrations of analyte and identical signals. If both experience the same loss of solvent their concentrations of analyte and signals will continue to be identical. In effect, we can ignore changes in concentration due to evaporation provided that the samples and standards experience an equivalent loss of solvent. If an identical standard and sample experience different losses of solvent,

however, their concentrations and signals will no longer be equal. In this case, an external standardization or standard addition results in a determinate error.

A standardization is still possible if the analyte's signal is referenced to a signal generated by another species that has been added at a fixed concentration to all samples and standards. The added species, which must be different from the analyte, is called an **internal standard.**

internal standard
A standard, whose identity is different from the analyte's, that is added to all samples and standards containing the analyte.

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

$$S_A = k_A C_A$$

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

where k_A and k_{IS} are the sensitivities for the analyte and internal standard, respectively. Taking the ratio of the two signals gives

$$\frac{S_A}{S_{IS}} = \frac{k_A}{k_{IS}} \times \frac{C_A}{C_{IS}} = K \times \frac{C_A}{C_{IS}} \qquad \textbf{5.10}$$

Because equation 5.10 is defined in terms of a ratio, K, of the analyte's sensitivity and the internal standard's sensitivity, it is not necessary to independently determine values for either k_A or k_{IS}.

In a single-point internal standardization, a single standard is prepared, and K is determined by solving equation 5.10

$$K = \left(\frac{C_{IS}}{C_A}\right)\left(\frac{S_A}{S_{IS}}\right)_{stand} \qquad \textbf{5.11}$$

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

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

Example

EXAMPLE 5.8

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

SOLUTION

Equation 5.11 allows us to calculate the value of K using the data for the standard

$$K = \left(\frac{C_{IS}}{C_A}\right)\left(\frac{S_A}{S_{IS}}\right)_{stand} = \frac{2.25}{1.75} \times 2.37 = 3.05$$

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

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

A single-point internal standardization has the same limitations as a single-point normal calibration. To construct an internal standard calibration curve, it is necessary to prepare several standards containing different concentrations of analyte. These standards are usually prepared such that the internal standard's concentration is constant. Under these conditions a calibration curve of $(S_A/S_{IS})_{stand}$ versus C_A is linear with a slope of K/C_{IS}.

Example

EXAMPLE 5.9

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

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

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

SOLUTION

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

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

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

When the internal standard's concentration cannot be held constant the data must be plotted as $(S_A/S_{IS})_{stand}$ versus C_A/C_{IS}, giving a linear calibration curve with a slope of K.

5C Linear Regression and Calibration Curves

In a single-point external standardization, we first determine the value of k by measuring the signal for a single standard containing a known concentration of analyte. This value of k and the signal for the sample are then used to calculate the concentration of analyte in the sample (see Example 5.2). With only a single determination of k, a quantitative analysis using a single-point external standardization is straightforward. This is also true for a single-point standard addition (see Examples 5.4 and 5.5) and a single-point internal standardization (see Example 5.8).

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

In a single-point standardization, we assume that the reagent blank (the first row in Table 5.1) corrects for all constant sources of determinate error. If this is not the case, then the value of k determined by a single-point standardization will have a determinate error.

Table 5.1 Data for Hypothetical Multiple-Point External Standardization

C_S	S_{meas}
0.000	0.00
0.100	12.36
0.200	24.83
0.300	35.91
0.400	48.79
0.500	60.42

Table 5.2 Effect of a Constant Determinate Error on the Value of *k* Calculated Using a Single-Point Standardization

C_A	S_{meas} (true)	*k* (true)	S_{meas} (with constant error)	*k* (apparent)
1.00	1.00	1.00	1.50	1.50
2.00	2.00	1.00	2.50	1.25
3.00	3.00	1.00	3.50	1.17
4.00	4.00	1.00	4.50	1.13
5.00	5.00	1.00	5.50	1.10
	mean *k*(true) =	1.00	mean *k* (apparent) =	1.23

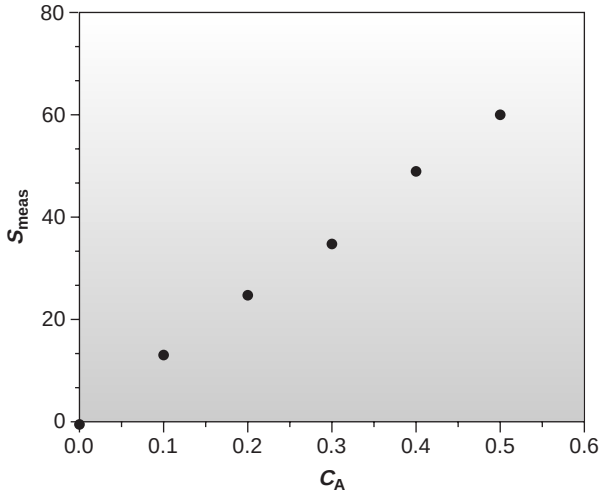


Figure 5.8
Normal calibration plot of hypothetical data from Table 5.1.

Table 5.2 demonstrates how an uncorrected constant error affects our determination of *k*. The first three columns show the concentration of analyte, the true measured signal (no constant error) and the true value of *k* for five standards. As expected, the value of *k* is the same for each standard. In the fourth column a constant determinate error of +0.50 has been added to the measured signals. The corresponding values of *k* are shown in the last column. Note that a different value of *k* is obtained for each standard and that all values are greater than the true value. As we noted in Section 5B.2, this is a significant limitation to any single-point standardization.

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

5C.1 Linear Regression of Straight-Line Calibration Curves

A calibration curve shows us the relationship between the measured signal and the analyte's concentration in a series of standards. The most useful calibration curve is a straight line since the method's sensitivity is the same for all concentrations of analyte. The equation for a linear calibration curve is

$$y = \beta_0 + \beta_1 x \qquad \textbf{5.12}$$

linear regression
A mathematical technique for fitting an equation, such as that for a straight line, to experimental data.

residual error
The difference between an experimental value and the value predicted by a regression equation.

where y is the signal and x is the amount of analyte. The constants β_0 and β_1 are the true y-intercept and the true slope, respectively. The goal of **linear regression** is to determine the best estimates for the slope, b_1, and y-intercept, b_0. This is accomplished by minimizing the **residual error** between the experimental values, y_i, and those values, $\hat{y}_i$, predicted by equation 5.12 (Figure 5.9). For obvious reasons, a regression analysis is also called a least-squares treatment. Several approaches to the linear regression of equation 5.12 are discussed in the following sections.

5C.2 Unweighted Linear Regression with Errors in *y*

The most commonly used form of linear regression is based on three assumptions: (1) that any difference between the experimental data and the calculated regression line is due to indeterminate errors affecting the values of *y*, (2) that these indeterminate errors are normally distributed, and (3) that the indeterminate errors in *y* do not depend on the value of *x*. Because we assume that indeterminate errors are the same for all standards, each standard contributes equally in estimating the slope and *y*-intercept. For this reason the result is considered an unweighted linear regression.

The second assumption is generally true because of the central limit theorem outlined in Chapter 4. The validity of the two remaining assumptions is less certain and should be evaluated before accepting the results of a linear regression. In particular, the first assumption is always suspect since there will certainly be some indeterminate errors affecting the values of *x*. In preparing a calibration curve, however, it is not unusual for the relative standard deviation of the measured signal (*y*) to be significantly larger than that for the concentration of analyte in the standards (*x*). In such circumstances, the first assumption is usually reasonable.

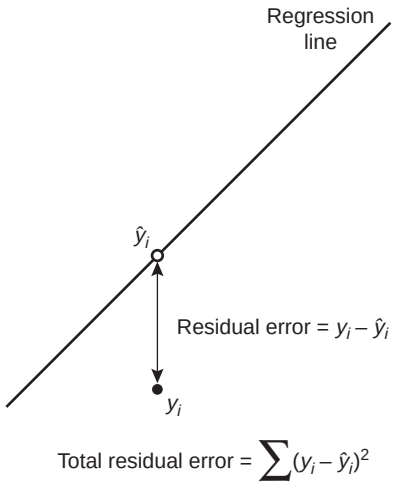


Figure 5.9
Residual error in linear regression, where the filled circle shows the experimental value y_i, and the open circle shows the predicted value $\hat{y}_i$.

Finding the Estimated Slope and *y*-Intercept The derivation of equations for calculating the estimated slope and *y*-intercept can be found in standard statistical texts[7] and is not developed here. The resulting equation for the slope is given as

$$b_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \qquad \textbf{5.13}$$

and the equation for the *y*-intercept is

$$b_0 = \frac{\sum y_i - b_1 \sum x_i}{n} \qquad \textbf{5.14}$$

Although equations 5.13 and 5.14 appear formidable, it is only necessary to evaluate four summation terms. In addition, many calculators, spreadsheets, and other computer software packages are capable of performing a linear regression analysis based on this model. To save time and to avoid tedious calculations, learn how to use one of these tools. For illustrative purposes, the necessary calculations are shown in detail in the following example.

Example

EXAMPLE 5.10

Using the data from Table 5.1, determine the relationship between S_{meas} and C_S by an unweighted linear regression.

SOLUTION

Equations 5.13 and 5.14 are written in terms of the general variables *x* and *y*. As you work through this example, remember that *x* represents the concentration of analyte in the standards (C_S), and that *y* corresponds to the signal (S_{meas}). We begin by setting up a table to help in the calculation of the summation terms Σx_i, Σy_i, Σx_i^2, and $\Sigma x_i y_i$ which are needed for the calculation of b_0 and b_1

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

Adding the values in each column gives

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

Substituting these values into equations 5.12 and 5.13 gives the estimated slope

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

and the estimated y-intercept

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

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

$$S_{meas} = 120.70 \times C_S + 0.21$$

Note that for now we keep enough significant figures to match the number of decimal places to which the signal was measured. The resulting calibration curve is shown in Figure 5.10.

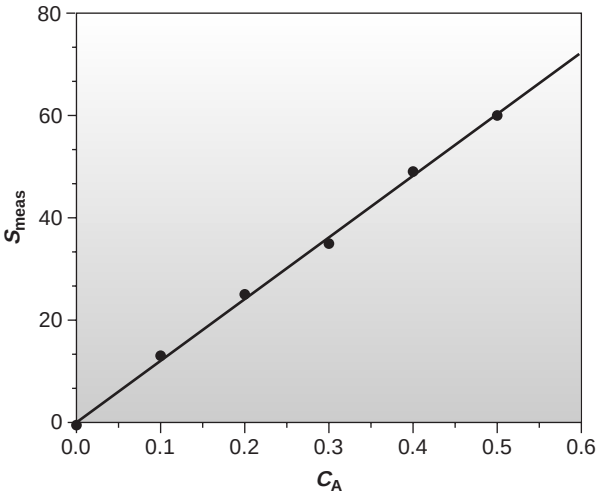


Figure 5.10
Normal calibration curve for the hypothetical data in Table 5.1, showing the regression line.

Uncertainty in the Regression Analysis As shown in Figure 5.10, the regression line need not pass through the data points (this is the consequence of indeterminate errors affecting the signal). The cumulative deviation of the data from the regression line is used to calculate the uncertainty in the regression due to

indeterminate error. This is called the **standard deviation about the regression,** s_r, and is given as

standard deviation about the regression The uncertainty in a regression analysis due to indeterminate error (s_r).

$$s_r = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} \qquad \textbf{5.15}$$

where y_i is the i^{th} experimental value, and $\hat{y}_i$ is the corresponding value predicted by the regression line

$$\hat{y}_i = b_0 + b_1 x_i$$

There is an obvious similarity between equation 5.15 and the standard deviation introduced in Chapter 4, except that the sum of squares term for s_r is determined relative to $\hat{y}_i$ instead of $\bar{y}$, and the denominator is $n-2$ instead of $n-1$; $n-2$ indicates that the linear regression analysis has only $n-2$ degrees of freedom since two parameters, the slope and the intercept, are used to calculate the values of $\hat{y}_i$.

A more useful representation of uncertainty is to consider the effect of indeterminate errors on the predicted slope and intercept. The standard deviation of the slope and intercept are given as

$$s_{b_1} = \sqrt{\frac{n s_r^2}{n\sum x_i^2 - (\sum x_i)^2}} = \sqrt{\frac{s_r^2}{\sum (x_i - \bar{x})^2}} \qquad \textbf{5.16}$$

$$s_{b_0} = \sqrt{\frac{s_r^2 \sum x_i^2}{n\sum x_i^2 - (\sum x_i)^2}} = \sqrt{\frac{s_r^2 \sum x_i^2}{n\sum (x_i - \bar{x})^2}} \qquad \textbf{5.17}$$

These standard deviations can be used to establish confidence intervals for the true slope and the true y-intercept

$$\beta_1 = b_1 \pm t s_{b_1} \qquad \textbf{5.18}$$

$$\beta_o = b_o \pm t s_{b_0} \qquad \textbf{5.19}$$

where t is selected for a significance level of α and for $n-2$ degrees of freedom. Note that the terms ts_{b_1} and ts_{b_0} do not contain a factor of $(\sqrt{n})^{-1}$ because the confidence interval is based on a single regression line. Again, many calculators, spreadsheets, and computer software packages can handle the calculation of s_{b_0} and s_{b_1} and the corresponding confidence intervals for β_0 and β_1. Example 5.11 illustrates the calculations.

Example

EXAMPLE 5.11

Calculate the 95% confidence intervals for the slope and y-intercept determined in Example 5.10.

SOLUTION

Again, as you work through this example, remember that x represents the concentration of analyte in the standards (C_S), and y corresponds to the signal (S_{meas}). To begin with, it is necessary to calculate the standard deviation about the regression. This requires that we first calculate the predicted signals, $\hat{y}_i$, using the slope and y-intercept determined in Example 5.10. Taking the first standard as an example, the predicted signal is

$$\hat{y}_i = b_0 + b_1 x = 0.209 + (120.706)(0.100) = 12.280$$

The results for all six solutions are shown in the following table.

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

Adding together the data in the last column gives the numerator of equation 5.15, $\Sigma(y_i - \hat{y}_i)^2$, as 0.6512. The standard deviation about the regression, therefore, is

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

Next we calculate s_{b_1} and s_{b_0} using equations 5.16 and 5.17. Values for the summation terms Σx_i^2 and Σx_i are found in Example 5.10.

$$s_{b_1} = \sqrt{\frac{n s_r^2}{n\sum x_i^2 - (\sum x_i)^2}} = \sqrt{\frac{(6)(0.4035)^2}{(6)(0.550) - (1.500)^2}} = 0.965$$

$$s_{b_0} = \sqrt{\frac{s_r^2 \sum x_i^2}{n\sum x_i^2 - (\sum x_i)^2}} = \sqrt{\frac{(0.4035)^2(0.550)}{(6)(0.550) - (1.500)^2}} = 0.292$$

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

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

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

The standard deviation about the regression, s_r, suggests that the measured signals are precise to only the first decimal place. For this reason, we report the slope and intercept to only a single decimal place.

To minimize the uncertainty in the predicted slope and y-intercept, calibration curves are best prepared by selecting standards that are evenly spaced over a wide range of concentrations or amounts of analyte. The reason for this can be rationalized by examining equations 5.16 and 5.17. For example, both s_{b_0} and s_{b_1} can be minimized by increasing the value of the term $\Sigma(x_i - \bar{x})^2$, which is present in the denominators of both equations. Thus, increasing the range of concentrations used in preparing standards decreases the uncertainty in the slope and the y-intercept. Furthermore, to minimize the uncertainty in the y-intercept, it also is necessary to decrease the value of the term Σx_i^2 in equation 5.17. This is accomplished by spreading the calibration standards evenly over their range.

Using the Regression Equation Once the regression equation is known, we can use it to determine the concentration of analyte in a sample. When using a normal calibration curve with external standards or an internal standards calibration curve, we measure an average signal for our sample, $\bar{Y}_X$, and use it to calculate the value of X

$$X = \frac{\overline{Y}_X - b_0}{b_1} \qquad \textbf{5.20}$$

The standard deviation for the calculated value of X is given by the following equation

$$s_X = \frac{s_r}{b_1}\left\{\frac{1}{m} + \frac{1}{n} + \frac{(\overline{Y}_X - \overline{y})^2}{b_1^2 \sum (x_i - \overline{x})^2}\right\}^{1/2} \qquad \textbf{5.21}$$

where m is the number of replicate samples used to establish $\overline{Y}_X$, n is the number of calibration standards, $\overline{y}$ is the average signal for the standards, and x_i and $\overline{x}$ are the individual and mean concentrations of the standards.[8] Once s_X is known the confidence interval for the analyte's concentration can be calculated as

$$\mu_X = X \pm ts_X$$

where μ_X is the expected value of X in the absence of determinate errors, and the value of t is determined by the desired level of confidence and for $n - 2$ degrees of freedom. The following example illustrates the use of these equations for an analysis using a normal calibration curve with external standards.

Example

EXAMPLE 5.12

Three replicate determinations are made of the signal for a sample containing an unknown concentration of analyte, yielding values of 29.32, 29.16, and 29.51. Using the regression line from Examples 5.10 and 5.11, determine the analyte's concentration, C_A, and its 95% confidence interval.

SOLUTION

The equation for a normal calibration curve using external standards is

$$S_{meas} = b_0 + b_1 \times C_A$$

thus, $\overline{Y}_X$ is the average signal of 29.33, and X is the analyte's concentration. Substituting the value of $\overline{Y}_X$ into equation 5.20 along with the estimated slope and the y-intercept for the regression line gives the analyte's concentration as

$$C_A = X = \frac{\overline{Y}_X - b_0}{b_1} = \frac{29.33 - 0.209}{120.706} = 0.241$$

To calculate the standard deviation for the analyte's concentration, we must determine the values for $\overline{y}$ and $\Sigma(x_i - \overline{x})^2$. The former is just the average signal for the standards used to construct the calibration curve. From the data in Table 5.1, we easily calculate that $\overline{y}$ is 30.385. Calculating $\Sigma(x_i - \overline{x})^2$ looks formidable, but we can simplify the calculation by recognizing that this sum of squares term is simply the numerator in a standard deviation equation; thus,

$$\sum (x_i - \overline{x})^2 = s^2(n - 1)$$

where s is the standard deviation for the concentration of analyte in the standards used to construct the calibration curve. Using the data in Table 5.1, we find that s is 0.1871 and

$$\sum (x_i - \overline{x})^2 = (0.1871)^2(6 - 1) = 0.175$$

Substituting known values into equation 5.21 gives

$$s_A = s_X = \frac{0.4035}{120.706}\left\{\frac{1}{3} + \frac{1}{6} + \frac{(29.33 - 30.385)^2}{(120.706)^2\,(0.175)}\right\}^{1/2} = 0.0024$$

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

$$\mu_A = C_A \pm ts_A = 0.241 \pm (2.78)(0.0024) = 0.241 \pm 0.007$$

In a standard addition the analyte's concentration is determined by extrapolating the calibration curve to find the x-intercept. In this case the value of X is

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

and the standard deviation in X is

$$s_X = \frac{s_r}{b_1}\left\{\frac{1}{n} + \frac{(\bar{y})^2}{b_1^2 \sum (x_i - \bar{x})^2}\right\}^{1/2}$$

where n is the number of standards used in preparing the standard additions calibration curve (including the sample with no added standard), and $\bar{y}$ is the average signal for the n standards. Because the analyte's concentration is determined by extrapolation, rather than by interpolation, s_X for the method of standard additions generally is larger than for a normal calibration curve.

A linear regression analysis should not be accepted without evaluating the validity of the model on which the calculations were based. Perhaps the simplest way to evaluate a regression analysis is to calculate and plot the residual error for each value of x. The residual error for a single calibration standard, r_i, is given as

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

If the regression model is valid, then the residual errors should be randomly distributed about an average residual error of 0, with no apparent trend toward either smaller or larger residual errors (Figure 5.11a). Trends such as those shown in Figures 5.11b and 5.11c provide evidence that at least one of the assumptions on which the regression model is based are incorrect. For example, the trend toward larger residual errors in Figure 5.11b suggests that the indeterminate errors affecting y are not independent of the value of x. In Figure 5.11c the residual errors are not randomly distributed, suggesting that the data cannot be modeled with a straight-line relationship. Regression methods for these two cases are discussed in the following sections.

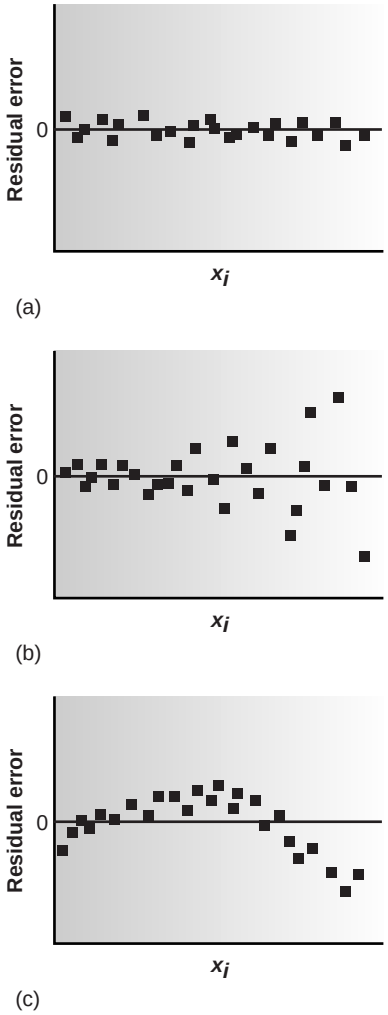


Figure 5.11
Plot of the residual error in y as a function of x. The distribution of the residuals in (a) indicates that the regression model was appropriate for the data, and the distributions in (b) and (c) indicate that the model does not provide a good fit for the data.

5C.3 Weighted Linear Regression with Errors in *y*

Equations 5.13 for the slope, b_1, and 5.14 for the y-intercept, b_0, assume that indeterminate errors equally affect each value of y. When this assumption is false, as shown in Figure 5.11b, the variance associated with each value of y must be included when estimating β_0 and β_1. In this case the predicted slope and intercept are

$$b_1 = \frac{n\sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{n\sum w_i x_i^2 - (\sum w_i x_i)^2} \qquad \textbf{5.22}$$

and

$$b_0 = \frac{\sum w_i y_i - b_1 \sum w_i x_i}{n} \qquad \textbf{5.23}$$

where w_i is a weighting factor accounting for the variance in measuring y_i. Values of w_i are calculated using equation 5.24.

$$w_i = \frac{n s_i^{-2}}{\sum s_i^{-2}} \qquad \textbf{5.24}$$

where s_i is the standard deviation associated with y_i. The use of a weighting factor ensures that the contribution of each pair of xy values to the regression line is proportional to the precision with which y_i is measured.

Example

EXAMPLE 5.13

The following data were recorded during the preparation of a calibration curve, where $\overline{S}_{meas}$ and s are the mean and standard deviation, respectively, for three replicate measurements of the signal.

C_A	$\overline{S}_{meas}$	s
0.000	0.00	0.02
0.100	12.36	0.02
0.200	24.83	0.07
0.300	35.91	0.13
0.400	48.79	0.22
0.500	60.42	0.33

Determine the relationship between $\overline{S}_{meas}$ and C_A using a weighted linear regression model.

SOLUTION

Once again, as you work through this example, remember that x represents the concentration of analyte in the standards (C_S), and y corresponds to the average signal ($\overline{S}_{meas}$). We begin by setting up a table to aid in the calculation of the weighting factor.

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

Adding together the values in the forth column gives

$$\sum s_i^{-2} = 5293.09$$

which is used to calculate the weights in the last column. As a check on the calculation, the sum of the weights in the last column should equal the number of calibration standards, n. In this case

$$\Sigma w_i = 6.0000$$

After the individual weights have been calculated, a second table is used to aid in calculating the four summation terms in equations 5.22 and 5.23.

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

Adding the values in the last four columns gives

$$\sum w_i x_i = 0.3644 \quad \sum w_i y_i = 44.9499 \quad \sum w_i x_i^2 = 0.0499 \quad \sum w_i x_i y_i = 6.1451$$

Substituting these values into the equations 5.22 and 5.23 gives the estimated slope

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

and the estimated y-intercept

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

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

$$\overline{S}_{meas} = 122.98 \times C_A + 0.02$$

with the calibration curve shown in Figure 5.12.

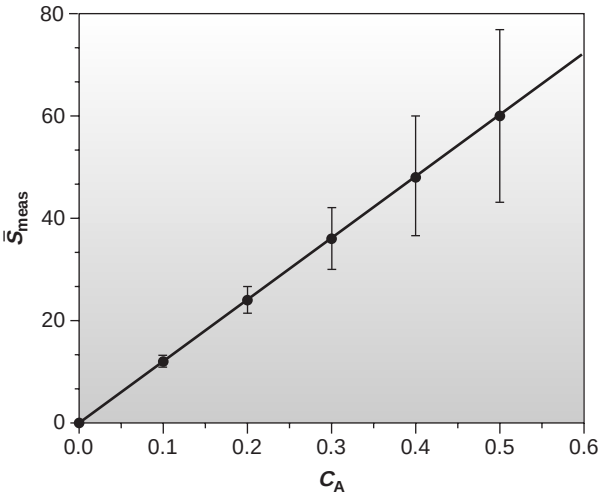


Figure 5.12
Weighted normal calibration curve for the data in Example 5.13. The lines through the data points show the standard deviation of the signal for the standards. These lines have been scaled by a factor of 50 so that they can be seen on the same scale as the calibration curve.

Equations for calculating confidence intervals for the slope, the *y*-intercept, and the concentration of analyte when using a weighted linear regression are not as easily defined as for an unweighted linear regression.[9] The confidence interval for the concentration of an analyte, however, will be at its optimum value when the analyte's signal is near the weighted centroid, $\bar{y}$, of the calibration curve

$$\bar{y} = \frac{1}{n}\sum w_i y_i$$

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

If we remove the assumption that the indeterminate errors affecting a calibration curve are found only in the signal (y), then indeterminate errors affecting the preparation of standards containing known amounts of analyte (x) must be factored into the regression model. The solution for the resulting regression line is computationally more involved than that for either the unweighted or weighted regression lines, and is not presented in this text. The suggested readings at the end of the chapter list several papers discussing algorithms for this regression method.

5C.5 Curvilinear and Multivariate Regression

Regression models based on a straight line, despite their apparent complexity, use the simplest functional relationship between two variables. In many cases, calibration curves show a pronounced curvature at high concentrations of analyte (see Figure 5.3b). One approach to constructing a calibration curve when curvature exists is to seek a transformation function that will make the data linear. Logarithms, exponentials, reciprocals, square roots, and trigonometric functions have all been used in this capacity. A plot of y versus log x is a typical example. Such transformations are not without complications. Perhaps the most obvious is that data that originally has a uniform variance for the y values will not maintain that uniform variance when the variable is transformed.

A more rigorous approach to developing a regression model for a nonlinear calibration curve is to fit a polynomial equation such as $y = a + bx + cx^2$ to the data. Equations for calculating the parameters *a, b,* and *c* are derived in the same manner as that described earlier for the straight-line model.[10] When a single polynomial equation cannot be fitted to the calibration data, it may be possible to fit separate polynomial equations to short segments of the calibration curve. The result is a single continuous calibration curve known as a spline function.

The regression models considered earlier apply only to functions containing a single independent variable. Analytical methods, however, are frequently subject to determinate sources of error due to interferents that contribute to the measured signal. In the presence of a single interferent, equations 5.1 and 5.2 become

$$S_{meas} = k_A n_A + k_I n_I + S_{reag}$$

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

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

5D Blank Corrections

In discussing ways to standardize a method, we assumed that an appropriate reagent blank had been used to correct S_{meas} for signals originating from sources other than the analyte. At that time we did not ask an important question—"What constitutes an appropriate reagent blank?" Surprisingly, the answer is not intuitively obvious.

In one study,[12] analytical chemists were asked to evaluate a data set consisting of a normal calibration curve, three samples of different size but drawn from the same source, and an analyte-free sample (Table 5.3). At least four different approaches for correcting the signals were used by the participants: (1) ignore the correction entirely, which clearly is incorrect; (2) use the *y*-intercept of the calibration curve as a calibration blank, CB; (3) use the analyte-free sample as a reagent blank, RB; and (4) use both the calibration and reagent blanks. Equations for calculating the concentration of analyte using each approach are shown in Table 5.4, along with the resulting concentration for the analyte in each of the three samples.

That all four methods give a different result for the concentration of analyte underscores the importance of choosing a proper blank but does not tell us which of the methods is correct. In fact, the variation within each method for the reported concentration of analyte indicates that none of these four methods has adequately corrected for the blank. Since the three samples were drawn from the same source, they must have the same true concentration of analyte. Since all four methods predict concentrations of analyte that are dependent on the size of the sample, we can conclude that none of these blank corrections has accounted for an underlying constant source of determinate error.

To correct for all constant method errors, a blank must account for signals due to the reagents and solvent used in the analysis and any bias due to interac-

Table 5.3 Hypothetical Data Used to Study Procedures for Method Blanks

W_s[a]	S_{stand}	Sample Number	W_x[b]	S_{samp}
1.6667	0.2500	1	62.4746	0.8000
5.0000	0.5000	2	82.7915	1.0000
8.3333	0.7500	3	103.1085	1.2000
9.5507	0.8413			
11.6667	1.0000		analyte-free[c]	0.1000
18.1600	1.4870			
19.9333	1.6200			

Calibration equation: $S_{stand} = 0.0750 \times W_s + 0.1250$

Source: Modified from Cardone, M. J. *Anal. Chem.* **1986,** *58,* 433–438.

[a] W_s = weight of analyte used to prepare standard solution by diluting to a fixed volume, *V*.

[b] W_x = weight of sample used to prepare sample solution by diluting to a fixed volume, *V*.

[c] Analyte-free sample prepared in the same fashion as samples, but without the analyte being present.

tions between the analyte and the sample matrix. Both the calibration blank and the reagent blank correct for signals due to the reagents and solvents. Any difference in their values is due to the number and composition of samples contributing to the determination of the blank.

Unfortunately, neither the calibration blank nor the reagent blank can correct for bias due to analyte–matrix interactions because the analyte is missing in the reagent blank, and the sample's matrix is missing from the calibration blank. The true method blank must include both the matrix and the analyte and, consequently, can only be determined using the sample itself. One approach is to measure the signal for samples of different size and determine the regression line from a plot of signal versus the amount of sample. The resulting y-intercept gives the signal for the condition of no sample and is known as the **total Youden blank.**[13] This is the true blank correction. The regression line for the sample data in Table 5.3 is

total Youden blank
A blank that corrects the signal for analyte–matrix interactions.

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

giving a true blank correction of 0.185. Using this value to correct the signals gives identical values for the concentration of analyte in all three samples (see Table 5.4, bottom row).

The total Youden blank is not encountered frequently in analytical work, because most chemists rely on a calibration blank when using calibration curves and rely on reagent blanks when using a single-point standardization. As long as any constant bias due to analyte–matrix interactions can be ignored, which is often the case, the accuracy of the method will not suffer. It is always a good idea, however, to check for constant sources of error, by analyzing samples of different sizes, before relying on either a calibration or reagent blank.

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

Approach for Correcting Method Blank	Equation[a]	Concentration of Analyte in Sample 1	Sample 2	Sample 3
Ignore blank corrections	$C_A = \frac{W_a}{W_x} = \frac{S_{samp}}{kW_x}$	0.1707	0.1610	0.1552
Use calibration blank	$C_A = \frac{W_a}{W_x} = \frac{S_{samp} - CB}{kW_x}$	0.1441	0.1409	0.1390
Use reagent blank	$C_A = \frac{W_a}{W_x} = \frac{S_{samp} - RB}{kW_x}$	0.1494	0.1449	0.1422
Use both calibration and reagent blank	$C_A = \frac{W_a}{W_x} = \frac{S_{samp} - CB - RB}{kW_x}$	0.1227	0.1248	0.1261
Use total Youden blank	$C_A = \frac{W_a}{W_x} = \frac{S_{samp} - TYB}{kW_x}$	0.1313	0.1313	0.1313

[a] C_A = concentration of analyte; W_a = weight of analyte; W_x = weight of sample; k = slope of calibration curve = 0.075 (see Table 5.3).
Abbreviations: CB = calibration blank = 0.125 (see Table 5.3); RB = reagent blank = 0.100 (see Table 5.3); TYB = total Youden blank = 0.185 (see text).

5E KEY TERMS

aliquot (*p. 111*)
external standard (*p. 109*)
internal standard (*p. 116*)
linear regression (*p. 118*)
matrix matching (*p. 110*)
method of standard additions (*p. 110*)
multiple-point standardization (*p. 109*)
normal calibration curve (*p. 109*)
primary reagent (*p. 106*)
reagent grade (*p. 107*)
residual error (*p. 118*)
secondary reagent (*p. 107*)
single-point standardization (*p. 108*)
standard deviation about the regression (*p. 121*)
total Youden blank (*p. 129*)

5F SUMMARY

In a quantitative analysis, we measure a signal and calculate the amount of analyte using one of the following equations.

$$S_{meas} = kn_A + S_{reag}$$

$$S_{meas} = kC_A + S_{reag}$$

To obtain accurate results we must eliminate determinate errors affecting the measured signal, S_{meas}, the method's sensitivity, k, and any signal due to the reagents, S_{reag}.

To ensure that S_{meas} is determined accurately, we calibrate the equipment or instrument used to obtain the signal. Balances are calibrated using standard weights. When necessary, we can also correct for the buoyancy of air. Volumetric glassware can be calibrated by measuring the mass of water contained or delivered and using the density of water to calculate the true volume. Most instruments have calibration standards suggested by the manufacturer.

An analytical method is standardized by determining its sensitivity. There are several approaches to standardization, including the use of external standards, the method of standard addition, and the use of an internal standard. The most desirable standardization strategy is an external standardization. The method of standard additions, in which known amounts of analyte are added to the sample, is used when the sample's matrix complicates the analysis. An internal standard, which is a species (not analyte) added to all samples and standards, is used when the procedure does not allow for the reproducible handling of samples and standards.

Standardizations using a single standard are common, but also are subject to greater uncertainty. Whenever possible, a multiple-point standardization is preferred. The results of a multiple-point standardization are graphed as a calibration curve. A linear regression analysis can provide an equation for the standardization.

A reagent blank corrects the measured signal for signals due to reagents other than the sample that are used in an analysis. The most common reagent blank is prepared by omitting the sample. When a simple reagent blank does not compensate for all constant sources of determinate error, other types of blanks, such as the total Youden blank, can be used.

5G *Suggested* EXPERIMENTS

Experiments

The following exercises and experiments help connect the material in this chapter to the analytical laboratory.

Calibration—Volumetric glassware (burets, pipets, and volumetric flasks) can be calibrated in the manner described in Example 5.1. Most instruments have a calibration sample that can be prepared to verify the instrument's accuracy and precision. For example, as described in this chapter, a solution of 60.06 ppm $K_2Cr_2O_7$ in 0.0050 M H_2SO_4 should give an absorbance of 0.640 ± 0.010 at a wavelength of 350.0 nm when using 0.0050 M H_2SO_4 as a reagent blank. These exercises also provide practice with using volumetric glassware, weighing samples, and preparing solutions.

Standardization—External standards, standard additions, and internal standards are a common feature of many quantitative analyses. Suggested experiments using these standardization methods are found in later chapters. A good project experiment for introducing external standardization, standard additions, and the importance of the sample's matrix is to explore the effect of pH on the quantitative analysis of an acid–base indicator. Using bromothymol blue as an example, external standards can be prepared in a pH 9 buffer and used to analyze samples buffered to different pHs in the range of 6–10. Results can be compared with those obtained using a standard addition.

5H PROBLEMS

1. In calibrating a 10-mL pipet, a measured volume of water was transferred to a tared flask and weighed, yielding a mass of 9.9814 g. (a) Calculate, with and without correcting for buoyancy, the volume of water delivered by the pipet. Assume that the density of water is 0.99707 g/cm^3 and that the density of the weights is 8.40 g/cm^3. (b) What are the absolute and relative errors introduced by failing to account for the effect of buoyancy? Is this a significant source of determinate error for the calibration of a pipet? Explain.

2. Repeat the questions in problem 1 for the case when a mass of 0.2500 g is measured for a solid that has a density of 2.50 g/cm^3.

3. Is the failure to correct for buoyancy a constant or proportional source of determinate error?

4. What is the minimum density of a substance necessary to keep the buoyancy correction to less than 0.01% when using brass calibration weights with a density of 8.40 g/cm^3?

5. Describe how you would use a serial dilution to prepare 100 mL each of a series of standards with concentrations of 1.000×10^{-5}, 1.000×10^{-4}, 1.000×10^{-3}, and 1.000×10^{-2} M from a 0.1000 M stock solution. Calculate the uncertainty for each solution using a propagation of uncertainty, and compare to the uncertainty if each solution was prepared by a single dilution of the stock solution. Tolerances for different types of volumetric glassware and digital pipets are found in Tables 4.2 and 4.4. Assume that the uncertainty in the molarity of the stock solution is ±0.0002.

6. Three replicate determinations of the signal for a standard solution of an analyte at a concentration of 10.0 ppm give values of 0.163, 0.157, and 0.161 (arbitrary units), respectively. The signal for a method blank was found to be 0.002. Calculate the concentration of analyte in a sample that gives a signal of 0.118.

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

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

9. An appropriate standard additions calibration curve based on equation 5.8 plots $S_{spike}(V_o + V_s)$ on the y-axis and C_sV_s on the x-axis. Clearly explain why you cannot plot S_{spike} on the y-axis and $C_s[V_s/(V_o + V_s)]$ on the x-axis. Derive equations for the slope and y-intercept, and explain how the amount of analyte in a sample can be determined from the calibration curve.

10. A standard sample was prepared containing 10.0 ppm of an analyte and 15.0 ppm of an internal standard. Analysis of the sample gave signals for the analyte and internal standard of 0.155 and 0.233 (arbitrary units), respectively. Sufficient internal standard was added to a sample to make it 15.0 ppm in the internal standard. Analysis of the sample yielded signals for the analyte and internal standard of 0.274 and 0.198, respectively. Report the concentration of analyte in the sample.

11. For each of the pairs of calibration curves in Figure 5.13 on page 132, select the calibration curve with the better set of standards. Briefly explain the reasons for your selections. The scales for the x-axes and y-axes are the same for each pair.

12. The following standardization data were provided for a series of external standards of Cd^{2+} that had been buffered to a pH of 4.6.[14]

$[Cd^{2+}]$ (nM)	15.4	30.4	44.9	59.0	72.7	86.0
S_{meas} (nA)	4.8	11.4	18.2	26.6	32.3	37.7

(a) Determine the standardization relationship by a linear regression analysis, and report the confidence intervals for the slope and y-intercept. (b) Construct a plot of the residuals, and comment on their significance.

At a pH of 3.7 the following data were recorded

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

(c) How much more or less sensitive is this method at the lower pH? (d) A single sample is buffered to a pH of 3.7 and analyzed for cadmium, yielding a signal of 66.3. Report the concentration of Cd^{2+} in the sample and its 95% confidence interval.

13. To determine the concentration of analyte in a sample, a standard additions was performed. A 5.00-mL portion of the sample was analyzed and then successive 0.10-mL spikes of a 600.0-ppb standard of the analyte

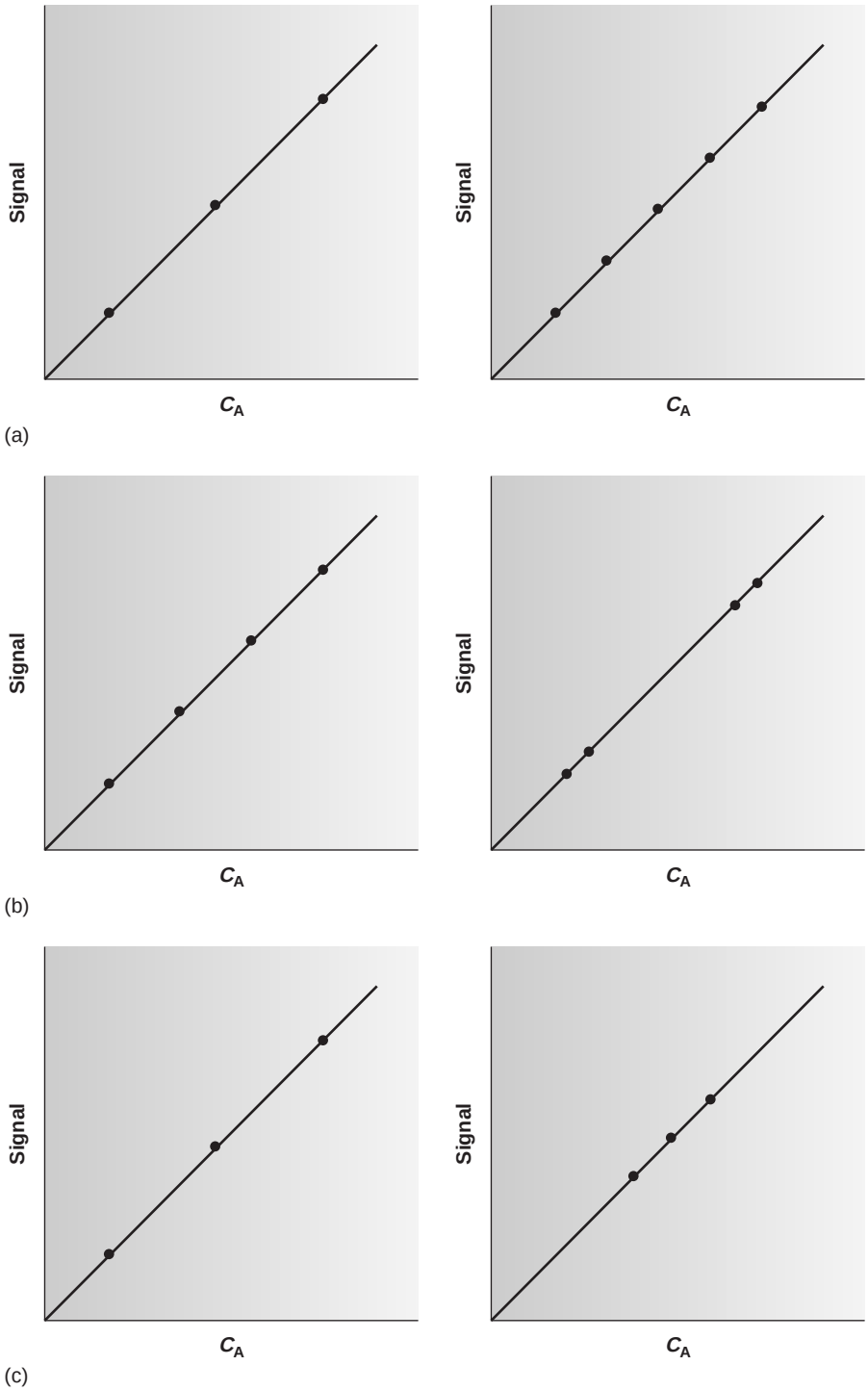

Signal
C_A
Signal
C_A
(a)
Signal
C_A
Signal
C_A
(b)
Signal
C_A
Signal
C_A
(c)

Figure 5.13

were added, analyzing after each spike. The following results were obtained

Volume of Spike (mL)	Signal (arbitrary units)
0.00	0.119
0.10	0.231
0.20	0.339
0.30	0.442

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

14. Troost and Olavesen investigated the application of an internal standardization to the quantitative analysis of polynuclear aromatic hydrocarbons.[15] The following results were obtained for the analysis of the analyte phenanthrene using isotopically labeled phenanthrene as an internal standard

	S_A/S_{IS}	
C_A/C_{IS}	Replicate 1	Replicate 2
0.50	0.514	0.522
1.25	0.993	1.024
2.00	1.486	1.471
3.00	2.044	2.080
4.00	2.342	2.550

(a) Determine the standardization relationship by a linear regression, and report the confidence intervals for the slope and *y*-intercept. (b) Based on your results, explain why the authors of this paper concluded that the internal standardization was inappropriate.

15. In Chapter 4 we used a paired *t*-test to compare two methods that had been used to independently analyze a series of samples of variable composition. An alternative approach is to plot the results for one method versus those for the other. If the two methods yield identical results, then the plot should have a true slope (β_1) of 1.00 and a true *y*-intercept (β_0) of 0.0. A *t*-test can be used to compare the actual slope and *y*-intercept with these ideal values. The appropriate test statistic for the *y*-intercept is found by rearranging equation 5.18

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

Rearranging equation 5.17 gives the test statistic for the slope

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

Reevaluate the data in problem 24 in Chapter 4 using the same significance level as in the original problem.*

16. Franke and co-workers evaluated a standard additions method for a voltammetric determination of Tl.[16] A summary of their results is tabulated here.

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

Determine the standardization relationship using a weighted linear regression.

5I SUGGESTED READINGS

In addition to the texts listed as suggested readings in Chapter 4, the following text provides additional details on regression

Draper, N. R.; Smith, H. *Applied Regression Analysis,* 2nd. ed. Wiley: New York, 1981.

Several articles providing more details about linear regression follow.

Boqué, R.; Rius, F. X.; Massart, D. L. "Straight Line Calibration: Something More Than Slopes, Intercepts, and Correlation Coefficients," *J. Chem. Educ.* **1993,** *70,* 230–232.

Henderson, G. "Lecture Graphic Aids for Least-Squares Analysis," *J. Chem. Educ.* **1988,** *65,* 1001–1003.

Renman, L., Jagner, D. "Asymmetric Distribution of Results in Calibration Curve and Standard Addition Evaluations," *Anal. Chim. Acta* **1997,** *357,* 157–166.

Two useful papers providing additional details on the method of standard additions are

Bader, M. "A Systematic Approach to Standard Addition Methods in Instrumental Analysis," *J. Chem. Educ.* **1980,** *57,* 703–706.

*Although this is a commonly used procedure for comparing two methods, it does violate one of the assumptions of an ordinary linear regression. Since both methods are expected to have indeterminate errors, an unweighted regression with errors in *y* may produce a biased result, with the slope being underestimated and the *y*-intercept being overestimated. This limitation can be minimized by placing the more precise method on the *x*-axis, using ten or more samples to increase the degrees of freedom in the analysis, and by using samples that uniformly cover the range of concentrations. For more information see Miller, J. C.; Miller, J. N. *Statistics for Analytical Chemistry,* 3rd ed. Ellis Horwood PTR Prentice-Hall: New York, 1993. Alternative approaches are discussed in Hartman, C.; Smeyers-Verbeke, J.; Penninckx, W.; Massart, D. L. *Anal. Chim. Acta* **1997,** *338,* 19–40 and Zwanziger, H. W.; Sârbu, C. *Anal. Chem.* **1998,** *70,* 1277–1280.

Nimura, Y.; Carr, M. R. "Reduction of the Relative Error in the Standard Additions Method," *Analyst* **1990,** *115,* 1589–1595.

The following paper discusses the importance of weighting experimental data when using linear regression

Karolczak, M. "To Weight or Not to Weight? An Analyst's Dilemma," *Curr. Separations* **1995,** *13,* 98–104.

Algorithms for performing a linear regression with errors in both x and y are discussed in

Irvin, J. A.; Quickenden, T. L. "Linear Least Squares Treatment When There Are Errors in Both x and y," *J. Chem. Educ.* **1983,** *60,* 711–712.

Kalantar, A. H. "Kerrich's Method for $y = \alpha x$ Data When Both y and x Are Uncertain," *J. Chem. Educ.* **1991,** *68,* 368–370.

Macdonald, J. R.; Thompson, W. J. "Least-Squares Fitting When Both Variables Contain Errors: Pitfalls and Possibilities," *Am. J. Phys.* **1992,** *60,* 66–73.

Ogren, P. J.; Norton, J. R. "Applying a Simple Linear Least-Squares Algorithm to Data with Uncertainties in Both Variables," *J. Chem. Educ.* **1992,** *69,* A130–A131.

The following paper discusses some of the problems that may be encountered when using linear regression to model data that have been mathematically transformed into a linear form.

Chong, D. P. "On the Use of Least Squares to Fit Data in Linear Form," *J. Chem. Educ.* **1994,** *71,* 489–490.

The analysis of nonlinear data is covered in the following papers.

Harris, D. C. "Nonlinear Least-Squares Curve Fitting with Microsoft Excel Solver," *J. Chem. Educ.* **1998,** *75,* 119–121.

Lieb, S. G. "Simplex Method of Nonlinear Least-Squares—A Logical Complementary Method to Linear Least-Squares Analysis of Data," *J. Chem. Educ.* **1997,** *74,* 1008–1011.

Machuca-Herrera, J. G. "Nonlinear Curve Fitting with Spreadsheets," *J. Chem. Educ.* **1997,** *74,* 448–449.

Zielinski, T. J.; Allendoerfer, R. D. "Least Squares Fitting of Nonlinear Data in the Undergraduate Laboratory," *J. Chem. Educ.* **1997,** *74,* 1001–1007.

More information on multivariate regression can be found in

Lang, P. M.; Kalivas, J. H. "A Global Perspective on Multivariate Calibration Methods," *J. Chemometrics* **1993,** *7,* 153–164.

Kowalski, B. R.; Seasholtz, M. B. "Recent Developments in Multivariate Calibration," *J. Chemometrics* **1991,** *5* 129–145.

An additional discussion on method blanks is found in the following two papers.

Cardone, M. J. "Detection and Determination of Error in Analytical Methodology. Part II. Correction for Corrigible Systematic Error in the Course of Real Sample Analysis," *J. Assoc. Off. Anal. Chem.* **1983,** *66,* 1283–1294.

Cardone, M. J. "Detection and Determination of Error in Analytical Methodology. Part IIB. Direct Calculational Technique for Making Corrigible Systematic Error Corrections," *J. Assoc. Off. Anal. Chem.* **1985,** *68,* 199–202.

5J REFERENCES

1. Battino, R.; Williamson, A. G. *J. Chem. Educ.* **1984,** *61,* 51–52.
2. Ebel, S. *Fresenius J. Anal. Chem.* **1992,** *342,* 769.
3. ACS Committee on Environmental Improvement "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," *Anal. Chem.* **1980,** *52,* 2242–2249.
4. Moody, J. R.; Greenburg, P. R.; Pratt, K. W.; et al. *Anal. Chem.* **1988,** *60,* 1203A–1218A.
5. Committee on Analytical Reagents, *Reagent Chemicals,* 8th ed., American Chemical Society: Washington, DC, 1993.
6. Cardone, M. J.; Palmero, P. J.; Sybrandt, L. B. *Anal. Chem.* **1980,** *52,* 1187–1191.
7. Draper, N. R.; Smith, H. *Applied Regression Analysis,* 2nd ed. Wiley: New York, 1981.
8. (a) Miller, J. N. *Analyst* **1991,** *116,* 3–14; and (b) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics,* Wiley-Interscience: New York, 1986; pp. 126–127.
9. Bonate, P. *J. Anal. Chem.* **1993,** *65,* 1367–1372.
10. (a) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics,* Wiley-Interscience: New York, 1986; (b) Deming, S. N.; Morgan, S. L. *Experimental Design: A Chemometric Approach,* Elsevier: Amsterdam, 1987.
11. Beebe, K. R.; Kowalski, B. R. *Anal. Chem.* **1987,** *59,* 1007A–1017A.
12. Cardone, M. *J. Anal. Chem.* **1986,** *58,* 433–438.
13. Cardone, M. *J. Anal. Chem.* **1986,** *58,* 438–445.
14. Wojciechowski, M; Balcerzak, *J. Anal. Chim. Acta* **1991,** *249,* 433–445.
15. Troost, J. R.; Olavesen, E. Y. *Anal. Chem.* **1996,** *68,* 708–711.
16. Franke, J. P.; de Zeeuw, R. A.; Hakkert, R. *Anal. Chem.* **1978,** *50,* 1374–1380.

Chapter 6

Equilibrium Chemistry

Regardless of the problem on which an analytical chemist is working, its solution ultimately requires a knowledge of chemistry and the ability to reason with that knowledge. For example, an analytical chemist developing a method for studying the effect of pollution on spruce trees needs to know, or know where to find, the structural and chemical differences between *p*-hydroxybenzoic acid and *p*-hydroxyacetophenone, two common phenols found in the needles of spruce trees (Figure 6.1). Chemical reasoning is a product of experience and is constructed from knowledge acquired in the classroom, the laboratory, and the chemical literature.

The material in this text assumes familiarity with topics covered in the courses and laboratory work you have already completed. This chapter provides a review of equilibrium chemistry. Much of the material in this chapter should be familiar to you, but other ideas are natural extensions of familiar topics.

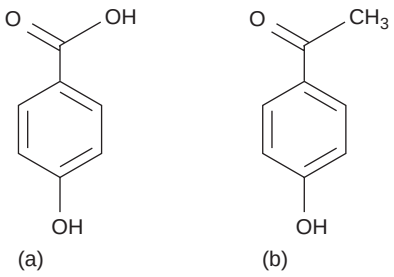


Figure 6.1
Structures of (a) *p*-hydroxybenzoic acid and (b) *p*-hydroxyacetophenone.

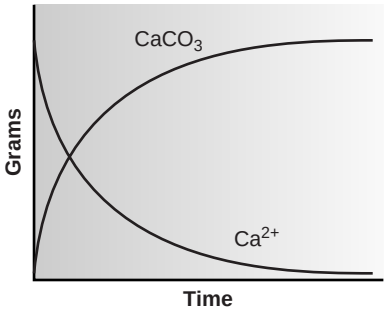


Figure 6.2
Change in mass of undissolved Ca^{2+} and solid $CaCO_3$ over time during the precipitation of $CaCO_3$.

equilibrium
A system is at equilibrium when the concentrations of reactants and products remain constant.

6A Reversible Reactions and Chemical Equilibria

In 1798, the chemist Claude Berthollet (1748–1822) accompanied a French military expedition to Egypt. While visiting the Natron Lakes, a series of salt water lakes carved from limestone, Berthollet made an observation that contributed to an important discovery. Upon analyzing water from the Natron Lakes, Berthollet found large quantities of common salt, NaCl, and soda ash, Na_2CO_3, a result he found surprising. Why would Berthollet find this result surprising and how did it contribute to an important discovery? Answering these questions provides an example of chemical reasoning and introduces the topic of this chapter.

Berthollet "knew" that a reaction between Na_2CO_3 and $CaCl_2$ goes to completion, forming NaCl and a precipitate of $CaCO_3$ as products.

$$Na_2CO_3 + CaCl_2 \rightarrow 2NaCl + CaCO_3$$

Understanding this, Berthollet expected that large quantities of NaCl and Na_2CO_3 could not coexist in the presence of $CaCO_3$. Since the reaction goes to completion, adding a large quantity of $CaCl_2$ to a solution of Na_2CO_3 should produce NaCl and $CaCO_3$, leaving behind no unreacted Na_2CO_3. In fact, this result is what he observed in the laboratory. The evidence from Natron Lakes, where the coexistence of NaCl and Na_2CO_3 suggests that the reaction has not gone to completion, ran counter to Berthollet's expectations. Berthollet's important insight was recognizing that the chemistry occurring in the Natron Lakes is the reverse of what occurs in the laboratory.

$$CaCO_3 + 2NaCl \rightarrow Na_2CO_3 + CaCl_2$$

Using this insight Berthollet reasoned that the reaction is reversible, and that the relative amounts of "reactants" and "products" determine the direction in which the reaction occurs, and the final composition of the reaction mixture. We recognize a reaction's ability to move in both directions by using a double arrow when writing the reaction.

$$Na_2CO_3 + CaCl_2 \rightleftharpoons 2NaCl + CaCO_3$$

Berthollet's reasoning that reactions are reversible was an important step in understanding chemical reactivity. When we mix together solutions of Na_2CO_3 and $CaCl_2$, they react to produce NaCl and $CaCO_3$. If we monitor the mass of dissolved Ca^{2+} remaining and the mass of $CaCO_3$ produced as a function of time, the result will look something like the graph in Figure 6.2. At the start of the reaction the mass of dissolved Ca^{2+} decreases and the mass of $CaCO_3$ increases. Eventually, however, the reaction reaches a point after which no further changes occur in the amounts of these species. Such a condition is called a state of **equilibrium.**

Although a system at equilibrium appears static on a macroscopic level, it is important to remember that the forward and reverse reactions still occur. A reaction at equilibrium exists in a "steady state," in which the rate at which any species forms equals the rate at which it is consumed.

6B Thermodynamics and Equilibrium Chemistry

Thermodynamics is the study of thermal, electrical, chemical, and mechanical forms of energy. The study of thermodynamics crosses many disciplines, including physics, engineering, and chemistry. Of the various branches of thermodynamics,

the most important to chemistry is the study of the changes in energy occurring during a chemical reaction.

Consider, for example, the general equilibrium reaction shown in equation 6.1, involving the solutes A, B, C, and D, with stoichiometric coefficients *a, b, c,* and *d.*

$$aA + bB \rightleftharpoons cC + dD \qquad \textbf{6.1}$$

By convention, species to the left of the arrows are called reactants, and those on the right side of the arrows are called products. As Berthollet discovered, writing a reaction in this fashion does not guarantee that the reaction of A and B to produce C and D is favorable. Depending on initial conditions, the reaction may move to the left, to the right, or be in a state of equilibrium. Understanding the factors that determine the final position of a reaction is one of the goals of chemical thermodynamics.

Chemical systems spontaneously react in a fashion that lowers their overall free energy. At a constant temperature and pressure, typical of many bench-top chemical reactions, the free energy of a chemical reaction is given by the **Gibb's free energy** function

Gibb's free energy
A thermodynamic function for systems at constant temperature and pressure that indicates whether or not a reaction is favorable ($\Delta G < 0$), unfavorable ($\Delta G > 0$), or at equilibrium ($\Delta G = 0$).

$$\Delta G = \Delta H - T\,\Delta S \qquad \textbf{6.2}$$

where T is the temperature in kelvins, and ΔG, ΔH, and ΔS are the differences in the Gibb's free energy, the enthalpy, and the entropy between the products and reactants.

Enthalpy is a measure of the net flow of energy, as heat, during a chemical reaction. Reactions in which heat is produced have a negative ΔH and are called exothermic. Endothermic reactions absorb heat from their surroundings and have a positive ΔH. **Entropy** is a measure of randomness, or disorder. The entropy of an individual species is always positive and tends to be larger for gases than for solids and for more complex rather than simpler molecules. Reactions that result in a large number of simple, gaseous products usually have a positive ΔS.

enthalpy
A change in enthalpy indicates the heat absorbed or released during a chemical reaction at constant pressure.

entropy
A measure of disorder.

The sign of ΔG can be used to predict the direction in which a reaction moves to reach its equilibrium position. A reaction is always thermodynamically favored when enthalpy decreases and entropy increases. Substituting the inequalities $\Delta H < 0$ and $\Delta S > 0$ into equation 6.2 shows that ΔG is negative when a reaction is thermodynamically favored. When ΔG is positive, the reaction is unfavorable as written (although the reverse reaction is favorable). Systems at equilibrium have a ΔG of zero.

As a system moves from a nonequilibrium to an equilibrium position, ΔG must change from its initial value to zero. At the same time, the species involved in the reaction undergo a change in their concentrations. The Gibb's free energy, therefore, must be a function of the concentrations of reactants and products.

As shown in equation 6.3, the Gibb's free energy can be divided into two terms.

$$\Delta G = \Delta G^\circ + RT \ln Q \qquad \textbf{6.3}$$

The first term, ΔG°, is the change in Gibb's free energy under **standard-state** conditions; defined as a temperature of 298 K, all gases with partial pressures of 1 atm, all solids and liquids pure, and all solutes present with 1 M concentrations. The second term, which includes the reaction quotient, Q, accounts for nonstandard-state pressures or concentrations. For reaction 6.1 the reaction quotient is

standard state
Condition in which solids and liquids are in pure form, gases have partial pressures of 1 atm, solutes have concentrations of 1 M, and the temperature is 298 K.

$$Q = \frac{[C]^c[D]^d}{[A]^a[B]^b} \qquad \textbf{6.4}$$

where the terms in brackets are the molar concentrations of the solutes. Note that the reaction quotient is defined such that the concentrations of products are placed

in the numerator, and the concentrations of reactants are placed in the denominator. In addition, each concentration term is raised to a power equal to its stoichiometric coefficient in the balanced chemical reaction. Partial pressures are substituted for concentrations when the reactant or product is a gas. The concentrations of pure solids and pure liquids do not change during a chemical reaction and are excluded from the reaction quotient.

At equilibrium the Gibb's free energy is zero, and equation 6.3 simplifies to

$$\Delta G^\circ = -RT \ln K$$

equilibrium constant
For a reaction at equilibrium, the equilibrium constant determines the relative concentrations of products and reactants.

where K is an **equilibrium constant** that defines the reaction's equilibrium position. The equilibrium constant is just the numerical value obtained when substituting the concentrations of reactants and products at equilibrium into equation 6.4; thus,

$$K = \frac{[C]_{eq}^{c}[D]_{eq}^{d}}{[A]_{eq}^{a}[B]_{eq}^{b}} \qquad \textbf{6.5}$$

where the subscript "eq" indicates a concentration at equilibrium. Although the subscript "eq" is usually omitted, it is important to remember that the value of K is determined by the concentrations of solutes at equilibrium.

As written, equation 6.5 is a limiting law that applies only to infinitely dilute solutions, in which the chemical behavior of any species in the system is unaffected by all other species. Corrections to equation 6.5 are possible and are discussed in more detail at the end of the chapter.

6C Manipulating Equilibrium Constants

We will use two useful relationships when working with equilibrium constants. First, if we reverse a reaction's direction, the equilibrium constant for the new reaction is simply the inverse of that for the original reaction. For example, the equilibrium constant for the reaction

$$A + 2B \rightleftharpoons AB_2 \qquad K_1 = \frac{[AB_2]}{[A][B]^2}$$

is the inverse of that for the reaction

$$AB_2 \rightleftharpoons A + 2B \qquad K_2 = \frac{1}{K_1} = \frac{[A][B]^2}{[AB_2]}$$

Second, if we add together two reactions to obtain a new reaction, the equilibrium constant for the new reaction is the product of the equilibrium constants for the original reactions.

$$A + C \rightleftharpoons AC \qquad K_1 = \frac{[AC]}{[A][C]}$$

$$AC + C \rightleftharpoons AC_2 \qquad K_2 = \frac{[AC_2]}{[AC][C]}$$

$$A + 2C \rightleftharpoons AC_2 \qquad K_3 = K_1K_2 = \frac{[AC]}{[A][C]} \times \frac{[AC_2]}{[AC][C]} = \frac{[AC_2]}{[A][C]^2}$$

Example

EXAMPLE 6.1

Calculate the equilibrium constant for the reaction

$$2A + B \rightleftharpoons C + 3D$$

given the following information

Rxn 1:	$A + B \rightleftharpoons D$	$K_1 = 0.40$
Rxn 2:	$A + E \rightleftharpoons C + D + F$	$K_2 = 0.10$
Rxn 3:	$C + E \rightleftharpoons B$	$K_3 = 2.0$
Rxn 4:	$F + C \rightleftharpoons D + B$	$K_4 = 5.0$

SOLUTION

The overall reaction is given as

$$\text{Rxn 1} + \text{Rxn 2} - \text{Rxn 3} + \text{Rxn 4}$$

If Rxn 3 is reversed, giving

$$\text{Rxn 5: } B \rightleftharpoons C + E \qquad K_5 = \frac{1}{K_3} = \frac{1}{2.0} = 0.50$$

then the overall reaction is

$$\text{Rxn 1} + \text{Rxn 2} + \text{Rxn 5} + \text{Rxn 4}$$

and the overall equilibrium constant is

$$K_{\text{overall}} = K_1 \times K_2 \times K_5 \times K_4 = 0.40 \times 0.10 \times 0.50 \times 5.0 = 0.10$$

6D Equilibrium Constants for Chemical Reactions

Several types of reactions are commonly used in analytical procedures, either in preparing samples for analysis or during the analysis itself. The most important of these are precipitation reactions, acid–base reactions, complexation reactions, and oxidation–reduction reactions. In this section we review these reactions and their equilibrium constant expressions.

6D.1 Precipitation Reactions

A precipitation reaction occurs when two or more soluble species combine to form an insoluble product that we call a **precipitate.** The most common precipitation reaction is a metathesis reaction, in which two soluble ionic compounds exchange parts. When a solution of lead nitrate is added to a solution of potassium chloride, for example, a precipitate of lead chloride forms. We usually write the balanced reaction as a net ionic equation, in which only the precipitate and those ions involved in the reaction are included. Thus, the precipitation of $PbCl_2$ is written as

precipitate
An insoluble solid that forms when two or more soluble reagents are combined.

$$Pb^{2+}(aq) + 2Cl^-(aq) \rightleftharpoons PbCl_2(s)$$

In the equilibrium treatment of precipitation, however, the reverse reaction describing the dissolution of the precipitate is more frequently encountered.

$$PbCl_2(s) \rightleftharpoons Pb^{2+}(aq) + 2Cl^-(aq)$$

solubility product
The equilibrium constant for a reaction in which a solid dissociates into its ions (K_{sp}).

The equilibrium constant for this reaction is called the **solubility product,** K_{sp}, and is given as

$$K_{sp} = [Pb^{2+}][Cl^-]^2 = 1.7 \times 10^{-5} \qquad \textbf{6.6}$$

Note that the precipitate, which is a solid, does not appear in the K_{sp} expression. It is important to remember, however, that equation 6.6 is valid only if $PbCl_2(s)$ is present and in equilibrium with the dissolved Pb^{2+} and Cl^-. Values for selected solubility products can be found in Appendix 3A.

6D.2 Acid–Base Reactions

acid
A proton donor.

base
A proton acceptor.

A useful definition of acids and bases is that independently introduced by Johannes Brønsted (1879–1947) and Thomas Lowry (1874–1936) in 1923. In the Brønsted-Lowry definition, **acids** are proton donors, and **bases** are proton acceptors. Note that these definitions are interrelated. Defining a base as a proton acceptor means an acid must be available to provide the proton. For example, in reaction 6.7 acetic acid, CH_3COOH, donates a proton to ammonia, NH_3, which serves as the base.

$$CH_3COOH(aq) + NH_3(aq) \rightleftharpoons CH_3COO^-(aq) + NH_4^+(aq) \qquad \textbf{6.7}$$

When an acid and a base react, the products are a new acid and base. For example, the acetate ion, CH_3COO^-, in reaction 6.7 is a base that reacts with the acidic ammonium ion, NH_4^+, to produce acetic acid and ammonia. We call the acetate ion the conjugate base of acetic acid, and the ammonium ion is the conjugate acid of ammonia.

Strong and Weak Acids The reaction of an acid with its solvent (typically water) is called an acid dissociation reaction. Acids are divided into two categories based on the ease with which they can donate protons to the solvent. Strong acids, such as HCl, almost completely transfer their protons to the solvent molecules.

$$HCl(aq) + H_2O(\ell) \rightarrow H_3O^+(aq) + Cl^-(aq)$$

In this reaction H_2O serves as the base. The hydronium ion, H_3O^+, is the conjugate acid of H_2O, and the chloride ion is the conjugate base of HCl. It is the hydronium ion that is the acidic species in solution, and its concentration determines the acidity of the resulting solution. We have chosen to use a single arrow ($\rightarrow$) in place of the double arrows ($\rightleftharpoons$) to indicate that we treat HCl as if it were completely dissociated in aqueous solutions. A solution of 0.10 M HCl is effectively 0.10 M in H_3O^+ and 0.10 M in Cl^-. In aqueous solutions, the common strong acids are hydrochloric acid (HCl), hydroiodic acid (HI), hydrobromic acid (HBr), nitric acid (HNO_3), perchloric acid ($HClO_4$), and the first proton of sulfuric acid (H_2SO_4).

Weak acids, of which aqueous acetic acid is one example, cannot completely donate their acidic protons to the solvent. Instead, most of the acid remains undissociated, with only a small fraction present as the conjugate base.

acid dissociation constant
The equilibrium constant for a reaction in which an acid donates a proton to the solvent (K_a).

$$CH_3COOH(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + CH_3COO^-(aq)$$

The equilibrium constant for this reaction is called an **acid dissociation constant,** K_a, and is written as

$$K_a = \frac{[H_3O^+][CH_3COO^-]}{[CH_3COOH]} = 1.75 \times 10^{-5}$$

Note that the concentration of H_2O is omitted from the K_a expression because its value is so large that it is unaffected by the dissociation reaction.* The magnitude of K_a provides information about the relative strength of a weak acid, with a smaller K_a corresponding to a weaker acid. The ammonium ion, for example, with a K_a of 5.70×10^{-10}, is a weaker acid than acetic acid.

Monoprotic weak acids, such as acetic acid, have only a single acidic proton and a single acid dissociation constant. Some acids, such as phosphoric acid, can donate more than one proton and are called polyprotic weak acids. Polyprotic acids are described by a series of acid dissociation steps, each characterized by it own acid dissociation constant. Phosphoric acid, for example, has three acid dissociation reactions and acid dissociation constants.

$$H_3PO_4(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + H_2PO_4^-(aq)$$

$$K_{a1} = \frac{[H_2PO_4^-][H_3O^+]}{[H_3PO_4]} = 7.11 \times 10^{-3}$$

$$H_2PO_4^-(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + HPO_4^{2-}(aq)$$

$$K_{a2} = \frac{[HPO_4^{2-}][H_3O^+]}{[H_2PO_4^-]} = 6.32 \times 10^{-8}$$

$$HPO_4^{2-}(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + PO_4^{3-}(aq)$$

$$K_{a3} = \frac{[PO_4^{3-}][H_3O^+]}{[HPO_4^{2-}]} = 4.5 \times 10^{-13}$$

The decrease in the acid dissociation constant from K_{a1} to K_{a3} tells us that each successive proton is harder to remove. Consequently, H_3PO_4 is a stronger acid than $H_2PO_4^-$, and $H_2PO_4^-$ is a stronger acid than HPO_4^{2-}.

Strong and Weak Bases Just as the acidity of an aqueous solution is a measure of the concentration of the hydronium ion, H_3O^+, the basicity of an aqueous solution is a measure of the concentration of the hydroxide ion, OH^-. The most common example of a strong base is an alkali metal hydroxide, such as sodium hydroxide, which completely dissociates to produce the hydroxide ion.

$$NaOH(aq) \rightarrow Na^+(aq) + OH^-(aq)$$

Weak bases only partially accept protons from the solvent and are characterized by a **base dissociation constant,** K_b. For example, the base dissociation reaction and base dissociation constant for the acetate ion are

base dissociation constant
The equilibrium constant for a reaction in which a base accepts a proton from the solvent (K_b).

$$CH_3COO^-(aq) + H_2O(\ell) \rightleftharpoons OH^-(aq) + CH_3COOH(aq)$$

$$K_b = \frac{[CH_3COOH][OH^-]}{[CH_3COO^-]} = 5.71 \times 10^{-10}$$

Polyprotic bases, like polyprotic acids, also have more than one base dissociation reaction and base dissociation constant.

Amphiprotic Species Some species can behave as either an acid or a base. For example, the following two reactions show the chemical reactivity of the bicarbonate ion, HCO_3^-, in water.

*The concentration of pure water is approximately 55.5 M

$$HCO_3^-(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + CO_3^{2-}(aq) \quad \textbf{6.8}$$

$$HCO_3^-(aq) + H_2O(\ell) \rightleftharpoons OH^-(aq) + H_2CO_3(aq) \quad \textbf{6.9}$$

amphiprotic
A species capable of acting as both an acid and a base.

A species that can serve as both a proton donor and a proton acceptor is called **amphiprotic.** Whether an amphiprotic species behaves as an acid or as a base depends on the equilibrium constants for the two competing reactions. For bicarbonate, the acid dissociation constant for reaction 6.8

$$K_{a2} = 4.69 \times 10^{-11}$$

is smaller than the base dissociation constant for reaction 6.9.

$$K_{b2} = 2.25 \times 10^{-8}$$

Since bicarbonate is a stronger base than it is an acid ($k_{b2} > k_{a2}$), we expect that aqueous solutions of HCO_3^- will be basic.

Dissociation of Water Water is an amphiprotic solvent in that it can serve as an acid or a base. An interesting feature of an amphiprotic solvent is that it is capable of reacting with itself as an acid and a base.

$$H_2O(\ell) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + OH^-(aq)$$

The equilibrium constant for this reaction is called water's dissociation constant, K_w,

$$K_w = [H_3O^+][OH^-] \quad \textbf{6.10}$$

which has a value of 1.0000×10^{-14} at a temperature of 24 °C. The value of K_w varies substantially with temperature. For example, at 20 °C, K_w is 6.809×10^{-15}, but at 30 °C K_w is 1.469×10^{-14}. At the standard state temperature of 25 °C, K_w is 1.008×10^{-14}, which is sufficiently close to 1.00×10^{-14} that the latter value can be used with negligible error.

The pH Scale An important consequence of equation 6.10 is that the concentrations of H_3O^+ and OH^- are related. If we know $[H_3O^+]$ for a solution, then $[OH^-]$ can be calculated using equation 6.10.

Example

EXAMPLE 6.2

What is the $[OH^-]$ if the $[H_3O^+]$ is 6.12×10^{-5} M?

SOLUTION

$$[OH^-] = \frac{K_w}{[H_3O^+]} = \frac{1.00 \times 10^{-14}}{6.12 \times 10^{-5}} = 1.63 \times 10^{-10}$$

pH
Defined as $pH = -\log[H_3O^+]$.

Equation 6.10 also allows us to develop a **pH** scale that indicates the acidity of a solution. When the concentrations of H_3O^+ and OH^- are equal, a solution is neither acidic nor basic; that is, the solution is neutral. Letting

$$[H_3O^+] = [OH^-]$$

and substituting into equation 6.10 leaves us with

$$K_w = [H_3O^+]^2 = 1.00 \times 10^{-14}$$

Solving for $[H_3O^+]$ gives

$$[H_3O^+] = \sqrt{1.00 \times 10^{-14}} = 1.00 \times 10^{-7}$$

A neutral solution has a hydronium ion concentration of 1.00×10^{-7} M and a pH of 7.00.* For a solution to be acidic, the concentration of H_3O^+ must be greater than that for OH^-, or

$$[H_3O^+] > 1.00 \times 10^{-7} \text{ M}$$

The pH of an acidic solution, therefore, must be less than 7.00. A basic solution, on the other hand, will have a pH greater than 7.00. Figure 6.3 shows the pH scale along with pH values for some representative solutions.

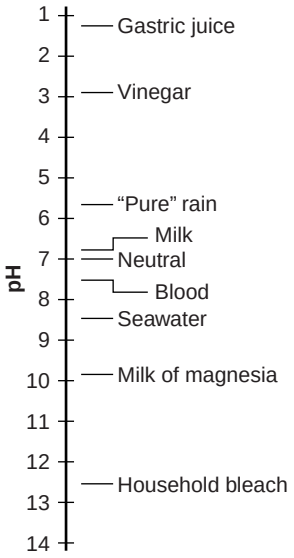


Figure 6.3
pH scale showing values for representative solutions.

Tabulating Values for K_a and K_b A useful observation about acids and bases is that the strength of a base is inversely proportional to the strength of its conjugate acid. Consider, for example, the dissociation reactions of acetic acid and acetate.

$$CH_3COOH(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + CH_3COO^-(aq) \qquad \textbf{6.11}$$

$$CH_3COO^-(aq) + H_2O(\ell) \rightleftharpoons CH_3COOH(aq) + OH^-(aq) \qquad \textbf{6.12}$$

Adding together these two reactions gives

$$2H_2O(\ell) \rightleftharpoons H_3O^+(aq) + OH^-(aq) \qquad \textbf{6.13}$$

The equilibrium constant for equation 6.13 is K_w. Since equation 6.13 is obtained by adding together reactions 6.11 and 6.12, K_w may also be expressed as the product of K_a for CH_3COOH and K_b for CH_3COO^-. Thus, for a weak acid, HA, and its conjugate weak base, A^-,

$$K_w = K_a \times K_b \qquad \textbf{6.14}$$

This relationship between K_a and K_b simplifies the tabulation of acid and base dissociation constants. Acid dissociation constants for a variety of weak acids are listed in Appendix 3B. The corresponding values of K_b for their conjugate weak bases are determined using equation 6.14.

Example

EXAMPLE 6.3

Using Appendix 3B, calculate the following equilibrium constants

(a) K_b for pyridine, C_5H_5N
(b) K_b for dihydrogen phosphate, $H_2PO_4^-$

SOLUTION

$$\text{(a)} \quad K_{b,C_5H_5N} = \frac{K_w}{K_{a,C_5H_5NH^+}} = \frac{1.00 \times 10^{-14}}{5.90 \times 10^{-6}} = 1.69 \times 10^{-9}$$

$$\text{(b)} \quad K_{b,H_2PO_4^-} = \frac{K_w}{K_{a,H_3PO_4}} = \frac{1.00 \times 10^{-14}}{7.11 \times 10^{-3}} = 1.41 \times 10^{-12}$$

*The use of a p-function to express a concentration is covered in Chapter 2.

6D.3 Complexation Reactions

A more general definition of acids and bases was proposed by G. N. Lewis (1875–1946) in 1923. The Brønsted–Lowry definition of acids and bases focuses on an acid's proton-donating ability and a base's proton-accepting ability. Lewis theory, on the other hand, uses the breaking and forming of covalent bonds to describe acid–base characteristics. In this treatment, an acid is an electron pair acceptor, and a base is an electron pair donor. Although Lewis theory can be applied to the treatment of acid–base reactions, it is more useful for treating complexation reactions between metal ions and ligands.

ligand
A Lewis base that binds with a metal ion.

The following reaction between the metal ion Cd^{2+} and the **ligand** NH_3 is typical of a complexation reaction.

$$Cd^{2+}(aq) + 4(:NH_3)(aq) \rightleftharpoons Cd(:NH_3)_4{}^{2+}(aq) \qquad \textbf{6.15}$$

The product of this reaction is called a metal–ligand complex. In writing the equation for this reaction, we have shown ammonia as $:NH_3$ to emphasize the pair of electrons it donates to Cd^{2+}. In subsequent reactions we will omit this notation.

formation constant
The equilibrium constant for a reaction in which a metal and a ligand bind to form a metal–ligand complex (K_f).

The formation of a metal–ligand complex is described by a **formation constant,** K_f. The complexation reaction between Cd^{2+} and NH_3, for example, has the following equilibrium constant

$$K_f = \frac{[Cd(NH_3)_4{}^{2+}]}{[Cd^{2+}][NH_3]^4} = 5.5 \times 10^7 \qquad \textbf{6.16}$$

dissociation constant
The equilibrium constant for a reaction in which a metal–ligand complex dissociates to form uncomplexed metal ion and ligand (K_d).

The reverse of reaction 6.15 is called a dissociation reaction and is characterized by a **dissociation constant,** K_d, which is the reciprocal of K_f.

Many complexation reactions occur in a stepwise fashion. For example, the reaction between Cd^{2+} and NH_3 involves four successive reactions

$$Cd^{2+}(aq) + NH_3(aq) \rightleftharpoons Cd(NH_3)^{2+}(aq) \qquad \textbf{6.17}$$

$$Cd(NH_3)^{2+}(aq) + NH_3(aq) \rightleftharpoons Cd(NH_3)_2{}^{2+}(aq) \qquad \textbf{6.18}$$

$$Cd(NH_3)_2{}^{2+}(aq) + NH_3(aq) \rightleftharpoons Cd(NH_3)_3{}^{2+}(aq) \qquad \textbf{6.19}$$

$$Cd(NH_3)_3{}^{2+}(aq) + NH_3(aq) \rightleftharpoons Cd(NH_3)_4{}^{2+}(aq) \qquad \textbf{6.20}$$

This creates a problem since it no longer is clear what reaction is described by a formation constant. To avoid ambiguity, formation constants are divided into two categories. **Stepwise formation constants,** which are designated as K_i for the ith step, describe the successive addition of a ligand to the metal–ligand complex formed in the previous step. Thus, the equilibrium constants for reactions 6.17–6.20 are, respectively, K_1, K_2, K_3, and K_4. Overall, or **cumulative formation constants,** which are designated as β_i, describe the addition of i ligands to the free metal ion. The equilibrium constant expression given in equation 6.16, therefore, is correctly identified as β_4, where

stepwise formation constant
The formation constant for a metal–ligand complex in which only one ligand is added to the metal ion or to a metal–ligand complex (K_i).

cumulative formation constant
The formation constant for a metal–ligand complex in which two or more ligands are simultaneously added to a metal ion or to a metal–ligand complex (β_i).

$$\beta_4 = K_1 \times K_2 \times K_3 \times K_4$$

In general

$$\beta_i = K_1 \times K_2 \times \cdots \times K_i$$

Stepwise and cumulative formation constants for selected metal–ligand complexes are given in Appendix 3C.

Equilibrium constants for complexation reactions involving solids are defined by combining appropriate K_{sp} and K_f expressions. For example, the solubility of AgCl increases in the presence of excess chloride as the result of the following complexation reaction

$$AgCl(s) + Cl^-(aq) \rightleftharpoons AgCl_2^-(aq) \qquad \textbf{6.21}$$

This reaction can be separated into three reactions for which equilibrium constants are known—the solubility of AgCl, described by its K_{sp}

$$AgCl(s) \rightleftharpoons Ag^+(aq) + Cl^-(aq)$$

and the stepwise formation of $AgCl_2^-$, described by K_1 and K_2

$$Ag^+(aq) + Cl^-(aq) \rightleftharpoons AgCl(aq)$$

$$AgCl(aq) + Cl^-(aq) \rightleftharpoons AgCl_2^-(aq)$$

The equilibrium constant for reaction 6.21, therefore, is equal to $K_{sp} \times K_1 \times K_2$.

Example

EXAMPLE 6.4

Determine the value of the equilibrium constant for the reaction

$$PbCl_2(s) \rightleftharpoons PbCl_2(aq)$$

SOLUTION

This reaction can be broken down into three reactions. The first of these reactions is the solubility of $PbCl_2$, described by its K_{sp}

$$PbCl_2(s) \rightleftharpoons Pb^{2+}(aq) + 2Cl^-(aq)$$

and the second and third are the stepwise formation of $PbCl_2\,(aq)$, described by K_1 and K_2

$$Pb^{2+}(aq) + Cl^-(aq) \rightleftharpoons PbCl^+(aq)$$

$$PbCl^+(aq) + Cl^-(aq) \rightleftharpoons PbCl_2(aq)$$

Using values for K_{sp}, K_1, and K_2 from Appendices 3A and 3C, we find the equilibrium constant to be

$$K = K_{sp} \times K_1 \times K_2 = (1.7 \times 10^{-5})(38.9)(1.62) = 1.1 \times 10^{-3}$$

6D.4 Oxidation–Reduction Reactions

In a complexation reaction, a Lewis base donates a pair of electrons to a Lewis acid. In an oxidation–reduction reaction, also known as a **redox reaction,** electrons are not shared, but are transferred from one reactant to another. As a result of this electron transfer, some of the elements involved in the reaction undergo a change in oxidation state. Those species experiencing an increase in their oxidation state are oxidized, while those experiencing a decrease in their oxidation state are reduced. For example, in the following redox reaction between Fe^{3+} and oxalic acid, $H_2C_2O_4$, iron is reduced since its oxidation state changes from +3 to +2.

redox reaction
An electron-transfer reaction.

$$2Fe^{3+}(aq) + H_2C_2O_4(aq) + 2H_2O(\ell) \rightleftharpoons 2Fe^{2+}(aq) + 2CO_2(g) + 2H_3O^+(aq) \qquad \textbf{6.22}$$

Oxalic acid, on the other hand, is oxidized since the oxidation state for carbon increases from +3 in $H_2C_2O_4$ to +4 in CO_2.

oxidation
A loss of electrons.

reduction
A gain of electrons.

Redox reactions, such as that shown in equation 6.22, can be divided into separate half-reactions that individually describe the **oxidation** and the **reduction** processes.

$$H_2C_2O_4(aq) + 2H_2O(\ell) \rightarrow 2CO_2(g) + 2H_3O^+(aq) + 2e^-$$

$$Fe^{3+}(aq) + e^- \rightarrow Fe^{2+}(aq)$$

It is important to remember, however, that oxidation and reduction reactions always occur in pairs.* This relationship is formalized by the convention of calling the species being oxidized a **reducing agent,** because it provides the electrons for the reduction half-reaction. Conversely, the species being reduced is called an **oxidizing agent.** Thus, in reaction 6.22, Fe^{3+} is the oxidizing agent and $H_2C_2O_4$ is the reducing agent.

reducing agent
A species that donates electrons to another species.

oxidizing agent
A species that accepts electrons from another species.

The products of a redox reaction also have redox properties. For example, the Fe^{2+} in reaction 6.22 can be oxidized to Fe^{3+}, while CO_2 can be reduced to $H_2C_2O_4$. Borrowing some terminology from acid–base chemistry, we call Fe^{2+} the conjugate reducing agent of the oxidizing agent Fe^{3+} and CO_2 the conjugate oxidizing agent of the reducing agent $H_2C_2O_4$.

Unlike the reactions that we have already considered, the equilibrium position of a redox reaction is rarely expressed by an equilibrium constant. Since redox reactions involve the transfer of electrons from a reducing agent to an oxidizing agent, it is convenient to consider the thermodynamics of the reaction in terms of the electron.

The free energy, ΔG, associated with moving a charge, Q, under a potential, E, is given by

$$\Delta G = EQ$$

Charge is proportional to the number of electrons that must be moved. For a reaction in which one mole of reactant is oxidized or reduced, the charge, in coulombs, is

$$Q = nF$$

where n is the number of moles of electrons per mole of reactant, and F is Faraday's constant (96,485 $C \cdot mol^{-1}$). The change in free energy (in joules per mole; J/mol) for a redox reaction, therefore, is

$$\Delta G = -nFE \qquad \textbf{6.23}$$

where ΔG has units of joules per mole. The appearance of a minus sign in equation 6.23 is due to a difference in the conventions for assigning the favored direction for reactions. In thermodynamics, reactions are favored when ΔG is negative, and redox reactions are favored when E is positive.

The relationship between electrochemical potential and the concentrations of reactants and products can be determined by substituting equation 6.23 into equation 6.3

$$-nFE = -nFE^\circ + RT \ln Q$$

where E° is the electrochemical potential under standard-state conditions. Dividing through by $-nF$ leads to the well-known **Nernst equation.**

Nernst equation
An equation relating electrochemical potential to the concentrations of products and reactants.

*Separating a redox reaction into its half-reactions is useful if you need to balance the reaction. One method for balancing redox reactions is reviewed in Appendix 4.

$$E = E^{o} - \frac{RT}{nF} \ln Q$$

Substituting appropriate values for R and F, assuming a temperature of 25 °C (298 K), and switching from ln to log* gives the potential in volts as

$$E = E^{o} - \frac{0.05916}{n} \log Q \qquad \textbf{6.24}$$

The standard-state electrochemical potential, E^{o}, provides an alternative way of expressing the equilibrium constant for a redox reaction. Since a reaction at equilibrium has a ΔG of zero, the electrochemical potential, E, also must be zero. Substituting into equation 6.24 and rearranging shows that

$$E^{o} = \frac{RT}{nF} \log K \qquad \textbf{6.25}$$

Standard-state potentials are generally not tabulated for chemical reactions, but are calculated using the standard-state potentials for the oxidation, E^{o}_{ox}, and reduction half-reactions, E^{o}_{red}. By convention, standard-state potentials are only listed for reduction half-reactions, and E^{o} for a reaction is calculated as

$$E^{o}_{reac} = E^{o}_{red} - E^{o}_{ox}$$

where both E^{o}_{red} and E^{o}_{ox} are standard-state reduction potentials.

Since the potential for a single half-reaction cannot be measured, a reference half-reaction is arbitrarily assigned a standard-state potential of zero. All other reduction potentials are reported relative to this reference. The standard half-reaction is

$$2H_3O^{+}(aq) + 2e^{-} \rightleftharpoons 2H_2O(\ell) + H_2(g)$$

Appendix 3D contains a listing of the standard-state reduction potentials for selected species. The more positive the standard-state reduction potential, the more favorable the reduction reaction will be under standard-state conditions. Thus, under standard-state conditions, the reduction of Cu^{2+} to Cu (E^{o} = +0.3419) is more favorable than the reduction of Zn^{2+} to Zn (E^{o} = –0.7618).

Example

EXAMPLE 6.5

Calculate (a) the standard-state potential, (b) the equilibrium constant, and (c) the potential when $[Ag^{+}]$ = 0.020 M and $[Cd^{2+}]$ = 0.050 M, for the following reaction taking place at 25 °C.

$$Cd(s) + 2Ag^{+}(aq) \rightleftharpoons Cd^{2+}(aq) + 2Ag(s)$$

SOLUTION

(a) In this reaction Cd is undergoing oxidation, and Ag^{+} is undergoing reduction. The standard-state cell potential, therefore, is

$$E^{o} = E^{o}_{Ag^{+}/Ag} - E^{o}_{Cd^{2+}/Cd} = 0.7996 \text{ V} - (-0.4030 \text{ V}) = 1.2026 \text{ V}$$

(b) To calculate the equilibrium constant, we substitute the values for the standard-state potential and number of electrons into equation 6.25.

$$1.2026 = \frac{0.05916}{2} \log K$$

*$\ln(x) = 2.303 \log(x)$

Solving for K gives the equilibrium constant as

$$\log K = 40.6558$$

$$K = 4.527 \times 10^{40}$$

(c) The potential when the $[Ag^+]$ is 0.020 M and the $[Cd^{2+}]$ is 0.050 M is calculated using equation 6.24 employing the appropriate relationship for the reaction quotient Q.

$$E = E^\circ - \frac{0.05916}{n} \log \frac{[Cd^{2+}]}{[Ag^+]^2}$$

$$= 1.2026 - \frac{0.05916}{2} \log \frac{(0.050)}{(0.020)^2}$$

$$= 1.14\text{V}$$

6E Le Châtelier's Principle

The equilibrium position for any reaction is defined by a fixed equilibrium constant, not by a fixed combination of concentrations for the reactants and products. This is easily appreciated by examining the equilibrium constant expression for the dissociation of acetic acid.

$$K_a = \frac{[H_3O^+][CH_3COO^-]}{[CH_3COOH]} = 1.75 \times 10^{-5} \qquad \textbf{6.26}$$

As a single equation with three variables, equation 6.26 does not have a unique solution for the concentrations of CH_3COOH, CH_3COO^-, and H_3O^+. At constant temperature, different solutions of acetic acid may have different values for $[H_3O^+]$, $[CH_3COO^-]$ and $[CH_3COOH]$, but will always have the same value of K_a.

If a solution of acetic acid at equilibrium is disturbed by adding sodium acetate, the $[CH_3COO^-]$ increases, suggesting an apparent increase in the value of K_a. Since K_a must remain constant, however, the concentration of all three species in equation 6.26 must change in a fashion that restores K_a to its original value. In this case, equilibrium is reestablished by the partial reaction of CH_3COO^- and H_3O^+ to produce additional CH_3COOH.

Le Châtelier's principle
When stressed, a system that was at equilibrium returns to its equilibrium state by reacting in a manner that relieves the stress.

The observation that a system at equilibrium responds to a stress by reequilibrating in a manner that diminishes the stress, is formalized as **Le Châtelier's principle.** One of the most common stresses that we can apply to a reaction at equilibrium is to change the concentration of a reactant or product. We already have seen, in the case of sodium acetate and acetic acid, that adding a product to a reaction mixture at equilibrium converts a portion of the products to reactants. In this instance, we disturb the equilibrium by adding a product, and the stress is diminished by partially reacting the excess product. Adding acetic acid has the opposite effect, partially converting the excess acetic acid to acetate.

In our first example, the stress to the equilibrium was applied directly. It is also possible to apply a concentration stress indirectly. Consider, for example, the following solubility equilibrium involving AgCl

$$AgCl(s) \rightleftharpoons Ag^+(aq) + Cl^-(aq) \qquad \textbf{6.27}$$

The effect on the solubility of AgCl of adding $AgNO_3$ is obvious,* but what is the effect of adding a ligand that forms a stable, soluble complex with Ag^+? Ammonia, for example, reacts with Ag^+ as follows

$$Ag^+(aq) + 2NH_3(aq) \rightleftharpoons Ag(NH_3)_2{}^+(aq) \qquad \textbf{6.28}$$

Adding ammonia decreases the concentration of Ag^+ as the $Ag(NH_3)_2{}^+$ complex forms. In turn, decreasing the concentration of Ag^+ increases the solubility of AgCl as reaction 6.27 reestablishes its equilibrium position. Adding together reactions 6.27 and 6.28 clarifies the effect of ammonia on the solubility of AgCl, by showing that ammonia is a reactant.

$$AgCl(s) + 2NH_3(aq) \rightleftharpoons Ag(NH_3)_2{}^+(aq) + Cl^-(aq) \qquad \textbf{6.29}$$

Example

EXAMPLE 6.6

What is the effect on the solubility of AgCl if HNO_3 is added to the equilibrium solution defined by reaction 6.29?

SOLUTION

Nitric acid is a strong acid that reacts with ammonia as shown here

$$HNO_3(aq) + NH_3(aq) \rightleftharpoons NH_4{}^+(aq) + NO_3{}^-(aq)$$

Adding nitric acid lowers the concentration of ammonia. Decreasing ammonia's concentration causes reaction 6.29 to move from products to reactants, decreasing the solubility of AgCl.

Increasing or decreasing the partial pressure of a gas is the same as increasing or decreasing its concentration.† The effect on a reaction's equilibrium position can be analyzed as described in the preceding example for aqueous solutes. Since the concentration of a gas depends on its partial pressure, and not on the total pressure of the system, adding or removing an inert gas has no effect on the equilibrium position of a gas-phase reaction.

Most reactions involve reactants and products that are dispersed in a solvent. If the amount of solvent is changed, either by diluting or concentrating the solution, the concentrations of all reactants and products either decrease or increase. The effect of these changes in concentration is not as intuitively obvious as when the concentration of a single reactant or product is changed. As an example, let's consider how dilution affects the equilibrium position for the formation of the aqueous silver-amine complex (reaction 6.28). The equilibrium constant for this reaction is

$$\beta_2 = \frac{[Ag(NH_3)_2{}^+]_{eq}}{[Ag^+]_{eq}[NH_3]^2_{eq}} \qquad \textbf{6.30}$$

*Adding $AgNO_3$ decreases the solubility of AgCl.

†The relationship between pressure and concentration can be deduced from the ideal gas law. Starting with $PV = nRT$, we solve for the molar concentration

$$\text{Molar concentration} = \frac{n}{V} = \frac{P}{RT}$$

Of course, this assumes an ideal gas (which is usually a reasonable assumption under normal laboratory conditions).

where the subscript "eq" is included for clarification. If a portion of this solution is diluted with an equal volume of water, each of the concentration terms in equation 6.30 is cut in half. Thus, the reaction quotient becomes

$$Q = \frac{(0.5)[Ag(NH_3)_2^+]_{eq}}{(0.5)[Ag^+]_{eq}(0.5)^2[NH_3]_{eq}^2}$$

which we can rewrite as

$$Q = \left(\frac{0.5}{(0.5)^3}\right)\left(\frac{[Ag(NH_3)_2^+]_{eq}}{[Ag^+][NH_3]_{eq}^2}\right) = 4 \times \beta_2$$

Since Q is greater than β_2, equilibrium must be reestablished by shifting the reaction to the left, decreasing the concentration of $Ag(NH_3)_2^+$. Furthermore, this new equilibrium position lies toward the side of the equilibrium reaction with the greatest number of solutes (one Ag^+ ion and two molecules of NH_3 versus the single metal–ligand complex). If the solution of $Ag(NH_3)_2^+$ is concentrated, by evaporating some of the solvent, equilibrium is reestablished in the opposite direction. This is a general conclusion that can be applied to any reaction, whether gas-phase, liquid-phase, or solid-phase. Increasing volume always favors the direction producing the greatest number of particles, and decreasing volume always favors the direction producing the fewest particles. If the number of particles is the same on both sides of the equilibrium, then the equilibrium position is unaffected by a change in volume.

6F Ladder Diagrams

When developing or evaluating an analytical method, we often need to understand how the chemistry taking place affects our results. We have already seen, for example, that adding NH_3 to a solution of Ag^+ is a poor idea if we intend to isolate the Ag^+ as a precipitate of AgCl (reaction 6.29). One of the primary sources of determinate method errors is a failure to account for potential chemical interferences.

ladder diagram
A visual tool for evaluating systems at equilibrium.

In this section we introduce the **ladder diagram** as a simple graphical tool for evaluating the chemistry taking place during an analysis.[1] Using ladder diagrams, we will be able to determine what reactions occur when several reagents are combined, estimate the approximate composition of a system at equilibrium, and evaluate how a change in solution conditions might affect our results.

6F.1 Ladder Diagrams for Acid–Base Equilibria

To see how a ladder diagram is constructed, we will use the acid–base equilibrium between HF and F^-

$$HF(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + F^-(aq)$$

for which the acid dissociation constant is

$$K_{a,HF} = \frac{[H_3O^+][F^-]}{[HF]}$$

Taking the log of both sides and multiplying through by –1 gives

$$-\log K_{a,HF} = -\log [H_3O^+] - \log \frac{[F^-]}{[HF]}$$

Finally, replacing the negative log terms with p-functions and rearranging leaves us with

$$\text{pH} = \text{p}K_\text{a} + \log\frac{[\text{F}^-]}{[\text{HF}]} \qquad \textbf{6.31}$$

Examining equation 6.31 tells us a great deal about the relationship between pH and the relative amounts of F^- and HF at equilibrium. If the concentrations of F^- and HF are equal, then equation 6.31 reduces to

$$\text{pH} = \text{p}K_{\text{a,HF}} = -\log(K_{\text{a,HF}}) = -\log(6.8 \times 10^{-4}) = 3.17$$

For concentrations of F^- greater than that of HF, the log term in equation 6.31 is positive and

$$\text{pH} > \text{p}K_{\text{a,HF}} \qquad \text{or} \qquad \text{pH} > 3.17$$

This is a reasonable result since we expect the concentration of hydrofluoric acid's conjugate base, F^-, to increase as the pH increases. Similar reasoning shows that the concentration of HF exceeds that of F^- when

$$\text{pH} < \text{p}K_{\text{a,HF}} \qquad \text{or} \qquad \text{pH} < 3.17$$

Now we are ready to construct the ladder diagram for HF (Figure 6.4). The ladder diagram consists of a vertical scale of pH values oriented so that smaller (more acidic) pH levels are at the bottom and larger (more basic) pH levels are at the top. A horizontal line is drawn at a pH equal to $\text{p}K_{\text{a,HF}}$. This line, or step, separates the solution into regions where each of the two conjugate forms of HF predominate. By referring to the ladder diagram, we see that at a pH of 2.5 hydrofluoric acid will exist predominately as HF. If we add sufficient base to the solution such that the pH increases to 4.5, the predominate form becomes F^-.

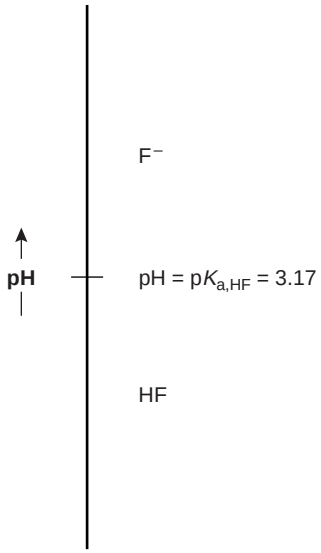


Figure 6.4
Ladder diagram for HF, showing areas of predominance for HF and F^-.

Figure 6.5 shows a second ladder diagram containing information about HF/F^- and NH_4^+/NH_3. From this ladder diagram we see that if the pH is less than 3.17, the predominate species are HF and NH_4^+. For pH's between 3.17 and 9.24 the predominate species are F^- and NH_4^+, whereas above a pH of 9.24 the predominate species are F^- and NH_3.

Ladder diagrams are particularly useful for evaluating the reactivity of acids and bases. An acid and a base cannot coexist if their respective areas of predominance do not overlap. If we mix together solutions of NH_3 and HF, the reaction

$$\text{HF}(aq) + \text{NH}_3(aq) \rightleftharpoons \text{NH}_4^+(aq) + \text{F}^-(aq) \qquad \textbf{6.32}$$

occurs because the predominance areas for HF and NH_3 do not overlap. Before continuing, let us show that this conclusion is reasonable by calculating the equilibrium constant for reaction 6.32. To do so we need the following three reactions and their equilibrium constants.

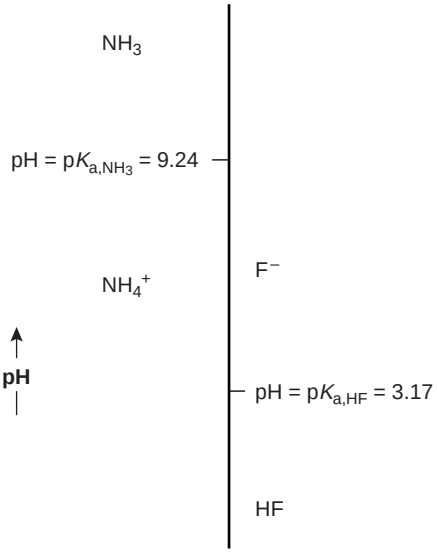


Figure 6.5
Ladder diagram for HF and NH_3.

$$\text{HF}(aq) + \text{H}_2\text{O}(\ell) \rightleftharpoons \text{H}_3\text{O}^+(aq) + \text{F}^-(aq) \qquad K_\text{a} = 6.8 \times 10^{-4}$$

$$\text{NH}_3(aq) + \text{H}_2\text{O}(\ell) \rightleftharpoons \text{OH}^-(aq) + \text{NH}_4^+(aq) \qquad K_\text{b} = 1.75 \times 10^{-5}$$

$$\text{H}_3\text{O}^+(aq) + \text{OH}^-(aq) \rightleftharpoons 2\text{H}_2\text{O}(\ell) \qquad K = \frac{1}{K_\text{w}} = \frac{1}{1.00 \times 10^{-14}}$$

Adding together these reactions gives us reaction 6.32, for which the equilibrium constant is

$$K = \frac{K_a K_b}{K_w} = \frac{(6.8 \times 10^{-4})(1.75 \times 10^{-5})}{(1.00 \times 10^{-14})} = 1.19 \times 10^{6}$$

Since the equilibrium constant is significantly greater than 1, the reaction's equilibrium position lies far to the right. This conclusion is general and applies to all ladder diagrams. The following example shows how we can use the ladder diagram in Figure 6.5 to evaluate the composition of any solution prepared by mixing together solutions of HF and NH_3.

Example

EXAMPLE 6.7

Predict the pH and composition of a solution prepared by adding 0.090 mol of HF to 0.040 mol of NH_3.

SOLUTION

Since HF is present in excess and the reaction between HF and NH_3 is favorable, the NH_3 will react to form NH_4^+. At equilibrium, essentially no NH_3 remains and

$$\text{Moles } NH_4^+ = 0.040 \text{ mol}$$

Converting NH_3 to NH_4^+ consumes 0.040 mol of HF; thus

$$\text{Moles HF} = 0.090 - 0.040 = 0.050 \text{ mol}$$

$$\text{Moles } F^- = 0.040 \text{ mol}$$

According to the ladder diagram for this system (see Figure 6.5), a pH of 3.17 results when there is an equal amount of HF and F^-. Since we have more HF than F^-, the pH will be slightly less than 3.17. Similar reasoning will show you that mixing together 0.090 mol of NH_3 and 0.040 mol of HF will result in a solution whose pH is slightly larger than 9.24.

If the areas of predominance for an acid and a base overlap each other, then practically no reaction occurs. For example, if we mix together solutions of NaF and NH_4Cl, we expect that there will be no significant change in the moles of F^- and NH_4^+. Furthermore, the pH of the mixture must be between 3.17 and 9.24. Because F^- and NH_4^+ can coexist over a range of pHs we cannot be more specific in estimating the solution's pH.

The ladder diagram for HF/F^- also can be used to evaluate the effect of pH on other equilibria that include either HF or F^-. For example, the solubility of CaF_2

$$CaF_2(s) \rightleftharpoons Ca^{2+}(aq) + 2F^-(aq)$$

is affected by pH because F^- is a weak base. Using Le Châtelier's principle, if F^- is converted to HF, the solubility of CaF_2 will increase. To minimize the solubility of CaF_2 we want to control the solution's pH so that F^- is the predominate species. From the ladder diagram we see that maintaining a pH of more than 3.17 ensures that solubility losses are minimal.

6F.2 Ladder Diagrams for Complexation Equilibria

The same principles used in constructing and interpreting ladder diagrams for acid–base equilibria can be applied to equilibria involving metal–ligand complexes. For complexation reactions the ladder diagram's scale is defined by the concentration of uncomplexed, or free ligand, pL. Using the formation of $Cd(NH_3)^{2+}$ as an example

$$Cd^{2+}(aq) + NH_3(aq) \rightleftharpoons Cd(NH_3)^{2+}(aq)$$

we can easily show that the dividing line between the predominance regions for Cd^{2+} and $Cd(NH_3)^{2+}$ is $\log(K_1)$.

$$K_1 = \frac{[Cd(NH_3)^{2+}]}{[Cd^{2+}][NH_3]}$$

$$\log(K_1) = \log\frac{[Cd(NH_3)^{2+}]}{[Cd^{2+}]} - \log[NH_3]$$

$$\log(K_1) = \log\frac{[Cd(NH_3)^{2+}]}{[Cd^{2+}]} + pNH_3$$

$$pNH_3 = \log(K_1) + \log\frac{[Cd^{2+}]}{[Cd(NH_3)^{2+}]}$$

Since K_1 for $Cd(NH_3)^{2+}$ is 3.55×10^2, $\log(K_1)$ is 2.55. Thus, for a pNH_3 greater than 2.55 (concentrations of NH_3 less than 2.8×10^{-3} M), Cd^{2+} is the predominate species. A complete ladder diagram for the metal–ligand complexes of Cd^{2+} and NH_3 is shown in Figure 6.6.

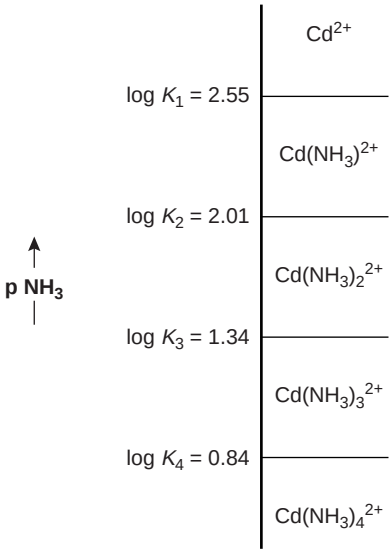


Figure 6.6
Ladder diagram for metal–ligand complexes of Cd^{2+} and NH_3.

Example

EXAMPLE 6.8

Using the ladder diagram in Figure 6.7, predict the result of adding 0.080 mol of Ca^{2+} to 0.060 mol of $Mg(EDTA)^{2-}$. EDTA is an abbreviation for the ligand ethylenediaminetetraacetic acid.

SOLUTION

The predominance regions for Ca^{2+} and $Mg(EDTA)^{2-}$ do not overlap, therefore, the reaction

$$Ca^{2+} + Mg(EDTA)^{2-} \rightleftharpoons Mg^{2+} + Ca(EDTA)^{2-}$$

will take place. Since there is an excess of Ca^{2+}, the composition of the final solution is approximately

$$\text{Moles } Ca^{2+} = 0.080 - 0.060 = 0.020 \text{ mol}$$

$$\text{Moles } Ca(EDTA)^{2-} = 0.060 \text{ mol}$$

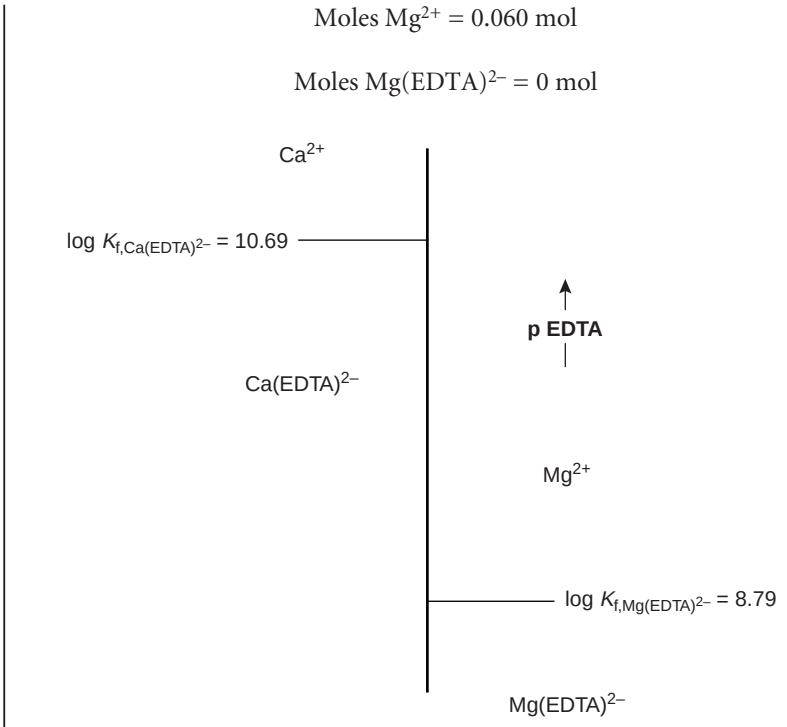


Figure 6.7
Ladder diagram for metal–ligand complexes of ethylenediaminetetraacetic acid (EDTA) with Ca^{2+} and Mg^{2+}.

We can also construct ladder diagrams using cumulative formation constants in place of stepwise formation constants. The first three stepwise formation constants for the reaction of Zn^{2+} with NH_3

$$Zn^{2+}(aq) + NH_3(aq) \rightleftharpoons Zn(NH_3)^{2+}(aq) \qquad K_1 = 1.6 \times 10^2$$

$$Zn(NH_3)^{2+}(aq) + NH_3(aq) \rightleftharpoons Zn(NH_3)_2{}^{2+}(aq) \qquad K_2 = 1.95 \times 10^2$$

$$Zn(NH_3)_2{}^{2+}(aq) + NH_3(aq) \rightleftharpoons Zn(NH_3)_3{}^{2+}(aq) \qquad K_3 = 2.3 \times 10^2$$

show that the formation of $Zn(NH_3)_3{}^{2+}$ is more favorable than the formation of $Zn(NH_3)^{2+}$ or $Zn(NH_3)_2{}^{2+}$. The equilibrium, therefore, is best represented by the cumulative formation reaction

$$Zn^{2+}(aq) + 3NH_3(aq) \rightleftharpoons Zn(NH_3)_3{}^{2+}(aq)$$

for which

$$\beta_3 = \frac{[Zn(NH_3)_3{}^{2+}]}{[Zn^{2+}][NH_3]^3} = 7.2 \times 10^6$$

Taking the log of each side gives

$$\log \beta_3 = \log \frac{[Zn(NH_3)_3{}^{2+}]}{[Zn^{2+}]} - 3 \log [NH_3]$$

or

$$pNH_3 = \frac{1}{3}\log\beta_3 + \frac{1}{3}\log\frac{[Zn^{2+}]}{[Zn(NH_3)_3{}^{2+}]}$$

The concentrations of Zn^{2+} and $Zn(NH_3)_3{}^{2+}$, therefore, are equal when

$$pNH_3 = \frac{1}{3}\log\beta_3 = \frac{1}{3}\log(7.2\times10^6) = 2.29$$

A complete ladder diagram for the Zn^{2+}–NH_3 system is shown in Figure 6.8.

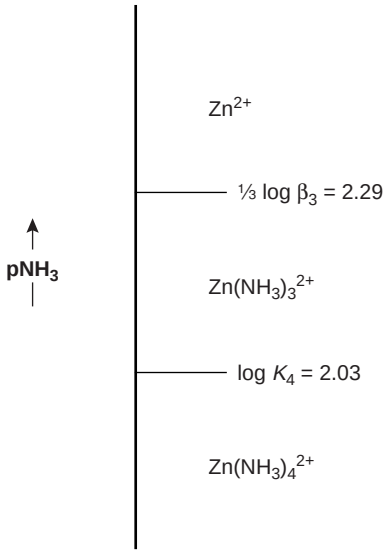


Figure 6.8
Ladder diagram for Zn^{2+}, $Zn(NH_3)_3{}^{2+}$, and $Zn(NH_3)_4{}^{2+}$, showing how cumulative formation constants are included.

6F.3 Ladder Diagram for Oxidation–Reduction Equilibria

Ladder diagrams can also be used to evaluate equilibrium reactions in redox systems. Figure 6.9 shows a typical ladder diagram for two half-reactions in which the scale is the electrochemical potential, *E*. Areas of predominance are defined by the Nernst equation. Using the Fe^{3+}/Fe^{2+} half-reaction as an example, we write

$$E = E^\circ - 0.05916\log\frac{[Fe^{2+}]}{[Fe^{3+}]} = +0.771\text{V} - 0.05916\log\frac{[Fe^{2+}]}{[Fe^{3+}]}$$

For potentials more positive than the standard-state potential, the predominate species is Fe^{3+}, whereas Fe^{2+} predominates for potentials more negative than E°. When coupled with the step for the Sn^{4+}/Sn^{2+} half-reaction, we see that Sn^{2+} can be used to reduce Fe^{3+}. If an excess of Sn^{2+} is added, the potential of the resulting solution will be near +0.154 V.

Using standard-state potentials to construct a ladder diagram can present problems if solutes are not at their standard-state concentrations. Because the concentrations of the reduced and oxidized species are in a logarithmic term, deviations from standard-state concentrations can usually be ignored if the steps being compared are separated by at least 0.3 V.[1b] A trickier problem occurs when a half-reaction's potential is affected by the concentration of another species. For example, the potential for the following half-reaction

$$UO_2{}^{2+}(aq) + 4H_3O^+(aq) + 2e^- \rightleftharpoons U^{4+}(aq) + 6H_2O(\ell)$$

depends on the pH of the solution. To define areas of predominance in this case, we begin with the Nernst equation

$$E = 0.327 - \frac{0.05916}{2}\log\frac{[U^{4+}]}{[UO_2{}^{2+}][H_3O^+]^4}$$

and factor out the concentration of H_3O^+.

$$E = 0.327 + \frac{0.05916}{2}\log[H_3O^+]^4 - \frac{0.05916}{2}\log\frac{[U^{4+}]}{[UO_2{}^{2+}]}$$

From this equation we see that the areas of predominance for $UO_2{}^{2+}$ and U^{4+} are defined by a step whose potential is

$$E = 0.327 + \frac{0.05916}{2}\log[H_3O^+]^4 = 0.327 - 0.1183\,\text{pH}$$

Figure 6.10 shows how a change in pH affects the step for the $UO_2{}^{2+}/U^{4+}$ half-reaction.

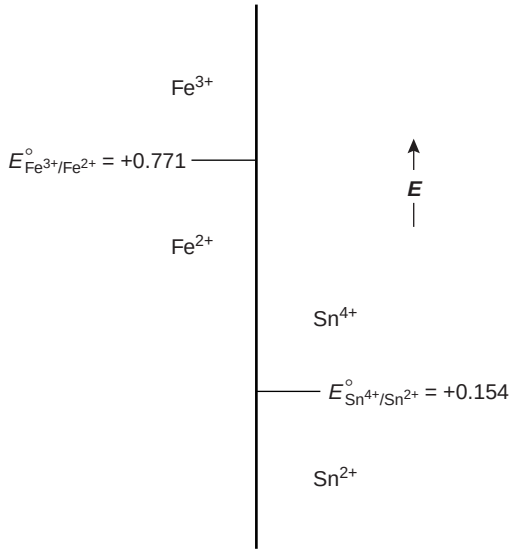


Figure 6.9
Ladder diagram for the Fe^{3+}/Fe^{2+} and Sn^{4+}/SN^{2+} half-reactions.

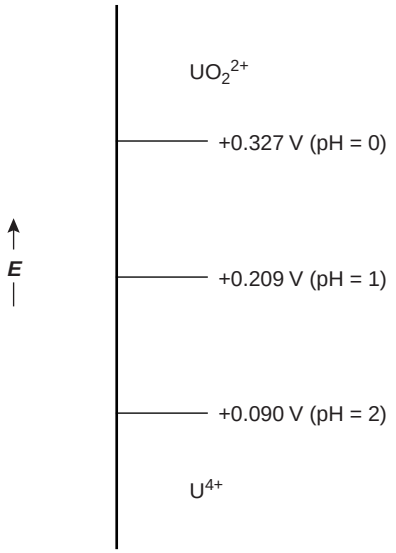


Figure 6.10
Ladder diagram showing the effect of a change in pH on the areas of predominance for the UO_2^{2+}/U^{4+} half-reaction.

6G Solving Equilibrium Problems

Ladder diagrams are a useful tool for evaluating chemical reactivity, usually providing a reasonable approximation of a chemical system's composition at equilibrium. When we need a more exact quantitative description of the equilibrium condition, a ladder diagram may not be sufficient. In this case we can find an algebraic solution. Perhaps you recall solving equilibrium problems in your earlier coursework in chemistry. In this section we will learn how to set up and solve equilibrium problems. We will start with a simple problem and work toward more complex ones.

6G.1 A Simple Problem: Solubility of $Pb(IO_3)_2$ in Water

When an insoluble compound such as $Pb(IO_3)_2$ is added to a solution a small portion of the solid dissolves. Equilibrium is achieved when the concentrations of Pb^{2+} and IO_3^- are sufficient to satisfy the solubility product for $Pb(IO_3)_2$. At equilibrium the solution is saturated with $Pb(IO_3)_2$. How can we determine the concentrations of Pb^{2+} and IO_3^-, and the solubility of $Pb(IO_3)_2$ in a saturated solution prepared by adding $Pb(IO_3)_2$ to distilled water?

We begin by writing the equilibrium reaction

$$Pb(IO_3)_2(s) \rightleftharpoons Pb^{2+}(aq) + 2IO_3^-(aq)$$

and its equilibrium constant

$$K_{sp} = [Pb^{2+}][IO_3^-]^2 = 2.5 \times 10^{-13} \qquad \textbf{6.33}$$

As equilibrium is established, two IO_3^- ions are produced for each ion of Pb^{2+}. If we assume that the molar concentration of Pb^{2+} at equilibrium is x then the molar concentration of IO_3^- is $2x$. To help keep track of these relationships, we can use the following table.

	$PbI_2(s)$	$\rightleftharpoons$	$Pb^{2+}(aq)$	+	$2IO_3^-(aq)$
Initial concentration	solid		0		0
Change in concentration	solid		$+x$		$+2x$
Equilibrium concentration	solid		$0 + x = x$		$0 + 2x = 2x$

Substituting the equilibrium concentrations into equation 6.33

$$(x)(2x)^2 = 2.5 \times 10^{-13}$$

and solving gives

$$4x^3 = 2.5 \times 10^{-13}$$

$$x = 3.97 \times 10^{-5}$$

The equilibrium concentrations of Pb^{2+} and IO_3^-, therefore, are

$$[Pb^{2+}] = x = 4.0 \times 10^{-5}\ M$$

$$[I^-] = 2x = 7.9 \times 10^{-5}\ M$$

Since one mole of $Pb(IO_3)_2$ contains one mole of Pb^{2+}, the solubility of $Pb(IO_3)_2$ is the same as the concentration of Pb^{2+}; thus, the solubility of $Pb(IO_3)_2$ is 4.0×10^{-5} M.

6G.2 A More Complex Problem: The Common Ion Effect

Calculating the solubility of $Pb(IO_3)_2$ in distilled water is a straightforward problem since the dissolution of the solid is the only source of Pb^{2+} or IO_3^-. How is the solubility of $Pb(IO_3)_2$ affected if we add $Pb(IO_3)_2$ to a solution of 0.10 M $Pb(NO_3)_2$? Before we set up and solve the problem algebraically, think about the chemistry occurring in this system, and decide whether the solubility of $Pb(IO_3)_2$ will increase, decrease, or remain the same. This is a good habit to develop. Knowing what answers are reasonable will help you spot errors in your calculations and give you more confidence that your solution to a problem is correct.

We begin by setting up a table to help us keep track of the concentrations of Pb^{2+} and IO_3^- in this system.

	$PbI_2(s)$	$\rightleftharpoons$	$Pb^{2+}(aq)$	+	$2IO_3^-(aq)$
Initial concentration	solid		0.10		0
Change in concentration	solid		$+x$		$+2x$
Equilibrium concentration	solid		$0.10 + x = x$		$0 + 2x = 2x$

Substituting the equilibrium concentrations into the solubility product expression (equation 6.33)

$$(0.10 + x)(2x)^2 = 2.5 \times 10^{-13}$$

and multiplying out the terms on the left leaves us with

$$4x^3 + 0.40x^2 = 2.5 \times 10^{-13} \qquad \textbf{6.34}$$

This is a more difficult equation to solve than that for the solubility of $Pb(IO_3)_2$ in distilled water, and its solution is not immediately obvious. A rigorous solution to equation 6.34 can be found using available computer software packages and spreadsheets.

How might we solve equation 6.34 if we do not have access to a computer? One possibility is that we can apply our understanding of chemistry to simplify the algebra. From Le Châtelier's principle, we expect that the large initial concentration of Pb^{2+} will significantly decrease the solubility of $Pb(IO_3)_2$. In this case we can reasonably expect the equilibrium concentration of Pb^{2+} to be very close to its initial concentration; thus, the following approximation for the equilibrium concentration of Pb^{2+} seems reasonable

$$[Pb^{2+}] = 0.10 + x \approx 0.10 \text{ M}$$

Substituting into equation 6.34

$$(0.10)(2x)^2 = 2.5 \times 10^{-13}$$

and solving for x gives

$$0.40x^2 = 2.5 \times 10^{-13}$$

$$x = 7.91 \times 10^{-7}$$

Before accepting this answer, we check to see if our approximation was reasonable. In this case the approximation $0.10 + x \approx 0.10$ seems reasonable since the difference between the two values is negligible. The equilibrium concentrations of Pb^{2+} and IO_3^-, therefore, are

$$[Pb^{2+}] = 0.10 + x \approx 0.10 \text{ M}$$

$$[I^-] = 2x = 1.6 \times 10^{-6} \text{ M}$$

common ion effect
The solubility of an insoluble salt decreases when it is placed in a solution already containing one of the salt's ions.

The solubility of $Pb(IO_3)_2$ is equal to the additional concentration of Pb^{2+} in solution, or 7.9×10^{-7} mol/L. As expected, the solubility of $Pb(IO_3)_2$ decreases in the presence of a solution that already contains one of its ions. This is known as the **common ion effect.**

As outlined in the following example, the process of making and evaluating approximations can be extended if the first approximation leads to an unacceptably large error.

Example

EXAMPLE 6.9

Calculate the solubility of $Pb(IO_3)_2$ in 1.0×10^{-4} M $Pb(NO_3)_2$.

SOLUTION

Letting x equal the change in the concentration of Pb^{2+}, the equilibrium concentrations are

$$[Pb^{2+}] = 1.0 \times 10^{-4} + x \qquad [IO_3^-] = 2x$$

and

$$(1.0 \times 10^{-4} + x)(2x)^2 = 2.5 \times 10^{-13}$$

We start by assuming that

$$[Pb^{2+}] = 1.0 \times 10^{-4} + x \approx 1.0 \times 10^{-4} \text{ M}$$

and solve for x, obtaining a value of 2.50×10^{-5}. Substituting back gives the calculated concentration of Pb^{2+} at equilibrium as

$$[Pb^{2+}] = 1.0 \times 10^{-4} + 2.50 \times 10^{-5} = 1.25 \times 10^{-4} \text{ M}$$

a value that differs by 25% from our approximation that the equilibrium concentration is 1.0×10^{-4} M. This error seems unreasonably large. Rather than shouting in frustration, we make a new assumption. Our first assumption that the concentration of Pb^{2+} is 1.0×10^{-4} M was too small. The calculated concentration of 1.25×10^{-4} M, therefore, is probably a little too large. Let us assume that

$$[Pb^{2+}] = 1.0 \times 10^{-4} + x \approx 1.2 \times 10^{-4} \text{ M}$$

Substituting into the solubility product equation and solving for x gives us

$$x = 2.28 \times 10^{-5}$$

or a concentration of Pb^{2+} at equilibrium of

$$[Pb^{2+}] = 1.0 \times 10^{-4} + (2.28 \times 10^{-5}) = 1.23 \times 10^{-4} \text{ M}$$

which differs from our assumed concentration of 1.2×10^{-4} M by 2.5%. This seems to be a reasonable error since the original concentration of Pb^{2+} is given to only two significant figures. Our final solution, to two significant figures, is

$$[Pb^{2+}] = 1.2 \times 10^{-4} \text{ M} \qquad [IO_3^-] = 4.6 \times 10^{-5} \text{ M}$$

and the solubility of $Pb(IO_3)_2$ is 2.3×10^{-5} mol/L. This iterative approach to solving an equation is known as the method of successive approximations.

6G.3 Systematic Approach to Solving Equilibrium Problems

Calculating the solubility of $Pb(IO_3)_2$ in a solution of $Pb(NO_3)_2$ was more complicated than calculating its solubility in distilled water. The necessary calculations, however, were still relatively easy to organize, and the assumption used to simplify the problem was fairly obvious. This problem was reasonably straightforward because it involved only a single equilibrium reaction, the solubility of $Pb(IO_3)_2$. Calculating the equilibrium composition of a system with multiple equilibrium reactions can become quite complicated. In this section we will learn how to use a systematic approach to setting up and solving equilibrium problems.

As its name implies, a systematic approach involves a series of steps:

1. Write all relevant equilibrium reactions and their equilibrium constant expressions.
2. Count the number of species whose concentrations appear in the equilibrium constant expressions; these are your unknowns. If the number of unknowns equals the number of equilibrium constant expressions, then you have enough information to solve the problem. If not, additional equations based on the conservation of mass and charge must be written. Continue to add equations until you have the same number of equations as you have unknowns.
3. Decide how accurate your final answer needs to be. This decision will influence your evaluation of any assumptions you use to simplify the problem.
4. Combine your equations to solve for one unknown (usually the one you are most interested in knowing). Whenever possible, simplify the algebra by making appropriate assumptions.
5. When you obtain your final answer, be sure to check your assumptions. If any of your assumptions prove invalid, then return to the previous step and continue solving. The problem is complete when you have an answer that does not violate any of your assumptions.

Besides equilibrium constant equations, two other types of equations are used in the systematic approach to solving equilibrium problems. The first of these is a **mass balance equation,** which is simply a statement of the conservation of matter. In a solution of a monoprotic weak acid, for example, the combined concentrations of the conjugate weak acid, HA, and the conjugate weak base, A^-, must equal the weak acid's initial concentration, C_{HA}.*

mass balance equation
An equation stating that matter is conserved, and that the total amount of a species added to a solution must equal the sum of the amount of each of its possible forms present in solution.

The second type of equation is a **charge balance equation.** A charge balance equation is a statement of solution electroneutrality.

Total positive charge from cations = total negative charge from anions

charge balance equation
An equation stating that the total concentration of positive charge in a solution must equal the total concentration of negative charge.

Mathematically, the charge balance expression is expressed as

$$\sum_{i=1}^{n} \left|(z^+)_i\right| \times [M^{z+}]_i = \sum_{j=1}^{m} \left|(z^-)_j\right| \times [A^{z-}]_j$$

where $[M^{z+}]_i$ and $[A^{z-}]_j$ are, respectively, the concentrations of the *i*th cation and the *j*th anion, and $(z^+)_i$ and $(z^-)_j$ are the charges of the *i*th cation and the *j*th anion. Note that the concentration terms are multiplied by the absolute values of each ion's charge, since electroneutrality is a conservation of charge, not concentration. Every ion in solution, even those not involved in any equilibrium

*You may recall that this is the difference between a formal concentration and a molar concentration.

reactions, must be included in the charge balance equation. The charge balance equation for an aqueous solution of $Ca(NO_3)_2$ is

$$2 \times [Ca^{2+}] + [H_3O^+] = [OH^-] + [NO_3^-]$$

Note that the concentration of Ca^{2+} is multiplied by 2, and that the concentrations of H_3O^+ and OH^- are also included. Charge balance equations must be written carefully since every ion in solution must be included. This presents a problem when the concentration of one ion in solution is held constant by a reagent of unspecified composition. For example, in many situations pH is held constant using a buffer. If the composition of the buffer is not specified, then a charge balance equation cannot be written.

Example

EXAMPLE 6.10

Write a mass balance and charge balance equations for a 0.10 M solution of $NaHCO_3$.

SOLUTION

It is easier to keep track of what species are in solution if we write down the reactions that control the solution's composition. These reactions are the dissolution of a soluble salt

$$NaHCO_3(s) \rightarrow Na^+(aq) + HCO_3^-(aq)$$

and the acid–base dissociation reactions of HCO_3^- and H_2O

$$HCO_3^-(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + CO_3^{2-}(aq)$$

$$HCO_3^-(aq) + H_2O(\ell) \rightleftharpoons OH^-(aq) + H_2CO_3(aq)$$

$$2H_2O(\ell) \rightleftharpoons H_3O^+(aq) + OH^-(aq)$$

The mass balance equations are

$$0.10\ M = [H_2CO_3] + [HCO_3^-] + [CO_3^{2-}]$$

$$0.10\ M = [Na^+]$$

The charge balance equation is

$$[Na^+] + [H_3O^+] = [OH^-] + [HCO_3^-] + 2 \times [CO_3^{2-}]$$

6G.4 pH of a Monoprotic Weak Acid

To illustrate the systematic approach, let us calculate the pH of 1.0 M HF. Two equilbria affect the pH of this system. The first, and most obvious, is the acid dissociation reaction for HF

$$HF(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + F^-(aq)$$

for which the equilibrium constant expression is

$$K_a = \frac{[H_3O^+][F^-]}{[HF]} = 6.8 \times 10^{-4} \qquad \textbf{6.35}$$

The second equilibrium reaction is the dissociation of water, which is an obvious yet easily disregarded reaction

$$2H_2O(\ell) \rightleftharpoons H_3O^+(aq) + OH^-(aq)$$

$$K_w = [H_3O^+][OH^-] = 1.00 \times 10^{-14} \quad \textbf{6.36}$$

Counting unknowns, we find four ([HF], $[F^-]$, $[H_3O^+]$, and $[OH^-]$). To solve this problem, therefore, we need to write two additional equations involving these unknowns. These equations are a mass balance equation

$$C_{HF} = [HF] + [F^-] \quad \textbf{6.37}$$

and a charge balance equation

$$[H_3O^+] = [F^-] + [OH^-] \quad \textbf{6.38}$$

We now have four equations (6.35, 6.36, 6.37, and 6.38) and four unknowns ([HF], $[F^-]$, $[H_3O^+]$, and $[OH^-]$) and are ready to solve the problem. Before doing so, however, we will simplify the algebra by making two reasonable assumptions. First, since HF is a weak acid, we expect the solution to be acidic; thus it is reasonable to assume that

$$[H_3O^+] >> [OH^-]$$

simplifying the charge balance equation (6.38) to

$$[H_3O^+] = [F^-] \quad \textbf{6.39}$$

Second, since HF is a weak acid we expect that very little dissociation occurs, and

$$[HF] >> [F^-]$$

Thus, the mass balance equation (6.36) simplifies to

$$C_{HF} = [HF] \quad \textbf{6.40}$$

For this exercise we will accept our assumptions if the error introduced by each assumption is less than ±5%.

Substituting equations 6.39 and 6.40 into the equilibrium constant expression for the dissociation of HF (equation 6.35) and solving for the concentration of H_3O^+ gives us

$$K_a = \frac{[H_3O^+][H_3O^+]}{C_{HF}}$$

$$[H_3O^+] = \sqrt{K_a C_{HF}} = \sqrt{(6.8 \times 10^{-4})(1.0)} = 2.6 \times 10^{-2}\ M$$

Before accepting this answer, we must verify that our assumptions are acceptable. The first assumption was that the $[OH^-]$ is significantly smaller than the $[H_3O^+]$. To calculate the concentration of OH^- we use the K_w expression (6.36)

$$[OH^-] = \frac{K_w}{[H_3O^+]} = \frac{1.00 \times 10^{-14}}{2.6 \times 10^{-2}} = 3.8 \times 10^{-13}\ M$$

Clearly this assumption is reasonable. The second assumption was that the $[F^-]$ is significantly smaller than the [HF]. From equation 6.39 we have

$$[F^-] = 2.6 \times 10^{-2}\ M$$

Since the $[F^-]$ is 2.6% of C_{HF}, this assumption is also within our limit that the error be no more than ±5%. Accepting our solution for the concentration of H_3O^+, we find that the pH of 1.0 M HF is 1.59.

How does the result of this calculation change if we require our assumptions to have an error of less than ±1%. In this case we can no longer assume that $[HF] >> [F^-]$. Solving the mass balance equation (6.37) for [HF]

$$[HF] = C_{HF} - [F^-]$$

and substituting into the K_a expression along with equation 6.39 gives

$$K_a = \frac{[H_3O^+]^2}{C_{HF} - [H_3O^+]}$$

Rearranging leaves us with a quadratic equation

$$[H_3O^+]^2 = K_aC_{HF} - K_a[H_3O^+]$$

$$[H_3O^+]^2 + K_a[H_3O^+] - K_aC_{HF} = 0$$

which we solve using the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

where *a, b,* and *c* are the coefficients in the quadratic equation $ax^2 + bx + c = 0$. Solving the quadratic formula gives two roots, only one of which has any chemical significance. For our problem the quadratic formula gives roots of

$$x = \frac{-6.8 \times 10^{-4} \pm \sqrt{(6.8 \times 10^{-4})^2 - (4)(1)(-6.8 \times 10^{-4})(1.0)}}{2(1)}$$

$$= \frac{-6.8 \times 10^{-4} \pm 5.22 \times 10^{-2}}{2}$$

$$= 2.57 \times 10^{-2} \quad \text{or} \quad -2.63 \times 10^{-2}$$

Only the positive root has any chemical significance since the negative root implies that the concentration of H_3O^+ is negative. Thus, the $[H_3O^+]$ is 2.6×10^{-2} M, and the pH to two significant figures is still 1.59.

This same approach can be extended to find the pH of a monoprotic weak base, replacing K_a with K_b, C_{HF} with the weak base's concentration, and solving for the $[OH^-]$ in place of $[H_3O^+]$.

Example

EXAMPLE 6.11

Calculate the pH of 0.050 M NH_3. State any assumptions made in simplifying the calculation, and verify that the error is less than 5%.

SOLUTION

Since NH_3 is a weak base ($K_b = 1.75 \times 10^{-5}$), we assume that

$$[OH^-] >> [H_3O^+] \qquad \text{and} \qquad C_{NH_3} = 0.050 \text{ M}$$

With these assumptions, we find (be sure to check the derivation)

$$[OH^-] = \sqrt{K_bC_{NH_3}} = \sqrt{(1.75 \times 10^{-5})(0.050)} = 9.35 \times 10^{-4} \text{ M}$$

Both assumptions are acceptable (again, verify that this is true). The concentration of H_3O^+ is calculated using K_w

$$[H_3O^+] = \frac{K_w}{[OH^-]} = \frac{1.00 \times 10^{-14}}{9.35 \times 10^{-4}} = 1.07 \times 10^{-11}$$

giving a pH of 10.97.

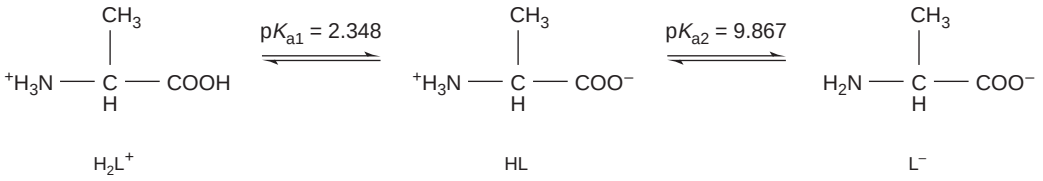


Figure 6.11
Acid–base equilibria for the amino acid alanine.

6G.5 pH of a Polyprotic Acid or Base

A more challenging problem is to find the pH of a solution prepared from a polyprotic acid or one of its conjugate species. As an example, we will use the amino acid alanine whose structure and acid dissociation constants are shown in Figure 6.11.

pH of 0.10 M H_2L^+ Alanine hydrochloride is a salt consisting of the diprotic weak acid H_2L^+ and Cl^-. Because H_2L^+ has two acid dissociation reactions, a complete systematic solution to this problem will be more complicated than that for a monoprotic weak acid. Using a ladder diagram (Figure 6.12) can help us simplify the problem. Since the areas of predominance for H_2L^+ and L^- are widely separated, we can assume that any solution containing an appreciable quantity of H_2L^+ will contain essentially no L^-. In this case, HL is such a weak acid that H_2L^+ behaves as if it were a monoprotic weak acid.

To find the pH of 0.10 M H_2L^+, we assume that

$$[H_3O^+] >> [OH^-]$$

Because H_2L^+ is a relatively strong weak acid, we cannot simplify the problem further, leaving us with

$$K_a = \frac{[H_3O^+]^2}{C_{H_2L^+} - [H_3O^+]}$$

Solving the resulting quadratic equation gives the $[H_3O^+]$ as 1.91×10^{-2} M or a pH of 1.72. Our assumption that $[H_3O^+]$ is significantly greater than $[OH^-]$ is acceptable.

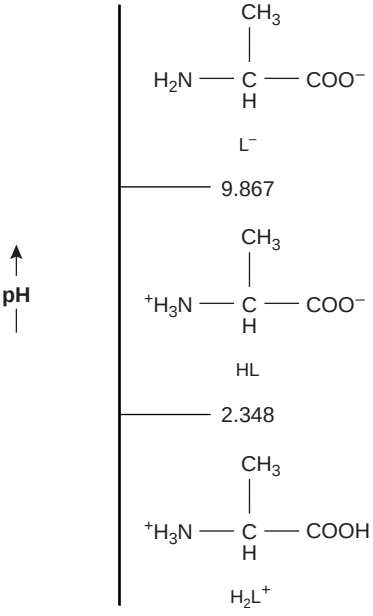


Figure 6.12
Ladder diagram for the amino acid alanine.

pH of 0.10 M L^- The alaninate ion is a diprotic weak base, but using the ladder diagram as a guide shows us that we can treat it as if it were a monoprotic weak base. Following the steps in Example 6.11 (which is left as an exercise), we find that the pH of 0.10 M alaninate is 11.42.

pH of 0.1 M HL Finding the pH of a solution of alanine is more complicated than that for H_2L^+ or L^- because we must consider two equilibrium reactions involving HL. Alanine is an amphiprotic species, behaving as an acid

$$HL(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + L^-(aq)$$

and a base

$$HL(aq) + H_2O(\ell) \rightleftharpoons OH^-(aq) + H_2L^+(aq)$$

As always, we must also consider the dissociation of water

$$2H_2O(\ell) \rightleftharpoons H_3O^+(aq) + OH^-(aq)$$

This leaves us with five unknowns ($[H_2L^+]$, [HL], $[L^-]$, $[H_3O^+]$, and $[OH^-]$), for which we need five equations. These equations are K_{a2} and K_{b2} for HL,

$$K_{a2} = \frac{[H_3O^+][L^-]}{[HL]}$$

$$K_{b2} = \frac{K_w}{K_{a1}} = \frac{[OH^-][H_2L^+]}{[HL]}$$

the K_w equation,

$$K_w = [H_3O^+][OH^-]$$

a mass balance equation on HL,

$$C_{HL} = [H_2L^+] + [HL] + [L^-]$$

and a charge balance equation

$$[H_2L^+] + [H_3O^+] = [OH^-] + [L^-]$$

From the ladder diagram it appears that we may safely assume that the concentrations of H_2L^+ and L^- are significantly smaller than that for HL, allowing us to simplify the mass balance equation to

$$C_{HL} = [HL]$$

Next we solve K_{b2} for $[H_2L^+]$

$$[H_2L^+] = \frac{K_w[HL]}{K_{a1}[OH^-]} = \frac{[HL][H_3O^+]}{K_{a1}} = \frac{C_{HL}[H_3O^+]}{K_{a1}}$$

and K_{a2} for $[L^-]$

$$[L^-] = \frac{K_{a2}[HL]}{[H_3O^+]} = \frac{K_{a2}C_{HL}}{[H_3O^+]}$$

Substituting these equations, along with the equation for K_w, into the charge balance equation gives us

$$\frac{C_{HL}[H_3O^+]}{K_{a1}} + [H_3O^+] = \frac{K_w}{[H_3O^+]} + \frac{K_{a2}C_{HL}}{[H_3O^+]}$$

which simplifies to

$$[H_3O^+]\left(\frac{C_{HL}}{K_{a1}} + 1\right) = \frac{1}{[H_3O^+]}(K_w + K_{a2}C_{HL})$$

$$[H_3O^+]^2 = \frac{K_{a2}C_{HL} + K_w}{(C_{HL}/K_{a1}) + 1}$$

$$[H_3O^+] = \sqrt{\frac{K_{a1}K_{a2}C_{HL} + K_{a1}K_w}{C_{HL} + K_{a1}}} \qquad \textbf{6.41}$$

We can simplify this equation further if $K_{a1}K_w << K_{a1}K_{a2}C_{HL}$, and if $K_{a1} << C_{HL}$, giving

$$[H_3O^+] = \sqrt{K_{a1}K_{a2}} \qquad \textbf{6.42}$$

For a solution of 0.10 M alanine, the $[H_3O^+]$ is

$$[H_3O^+] = \sqrt{(4.487 \times 10^{-3})(1.358 \times 10^{-10})} = 7.807 \times 10^{-7}\ \text{M}$$

or a pH of 6.11. Verifying that the assumptions are acceptable is left as an exercise.

Triprotic Acids and Bases, and Beyond The treatment of a diprotic acid or base is easily extended to acids and bases having three or more acid–base sites. For a triprotic weak acid such as H_3PO_4, for example, we can treat H_3PO_4 as if it was a monoprotic weak acid, $H_2PO_4^-$ and HPO_4^{2-} as if they were intermediate forms of diprotic weak acids, and PO_4^{3-} as if it was a monoprotic weak base.

Example

EXAMPLE 6.12

Calculate the pH of 0.10 M Na_2HPO_4.

SOLUTION

We treat HPO_4^{2-} as the intermediate form of a diprotic weak acid

$$H_2PO_4^-(aq) \rightleftharpoons HPO_4^{2-}(aq) \rightleftharpoons PO_4^{3-}(aq)$$

where the equilibrium constants are $K_{a2} = 6.32 \times 10^{-8}$ and $K_{a3} = 4.5 \times 10^{-13}$. Since the value of K_{a3} is so small, we use equation 6.41 instead of equation 6.42.

$$[H_3O^+] = \sqrt{\frac{(6.32 \times 10^{-8})(4.5 \times 10^{-13})(0.10) + (6.32 \times 10^{-8})(1.00 \times 10^{-14})}{0.10 + 6.32 \times 10^{-8}}}$$

$$= 1.86 \times 10^{-10}$$

or a pH of 9.73.

6G.6 Effect of Complexation on Solubility

The solubility of a precipitate can be improved by adding a ligand capable of forming a soluble complex with one of the precipitate's ions. For example, the solubility of AgI increases in the presence of NH_3 due to the formation of the soluble $Ag(NH_3)_2^+$ complex. As a final illustration of the systematic approach to solving equilibrium problems, let us find the solubility of AgI in 0.10 M NH_3.

We begin by writing the equilibria that we need to consider

$$AgI(s) \rightleftharpoons Ag^+(aq) + I^-(aq)$$

$$Ag^+(aq) + 2NH_3(aq) \rightleftharpoons Ag(NH_3)_2^+(aq)$$

$$NH_3(aq) + H_2O(\ell) \rightleftharpoons OH^-(aq) + NH_4^+(aq)$$

$$2H_2O(\ell) \rightleftharpoons H_3O^+(aq) + OH^-(aq)$$

Counting unknowns, we find that there are seven—$[Ag^+]$, $[I^-]$, $[Ag(NH_3)_2^+]$, $[NH_3]$, $[NH_4^+]$, $[OH^-]$, and $[H_3O^+]$. Four of the equations needed to solve this problem are given by the equilibrium constant expressions

$$K_{sp} = [Ag^+][I^-] = 8.3 \times 10^{-17}$$

$$\beta_2 = \frac{[Ag(NH_3)_2^+]}{[Ag^+][NH_3]^2} = 1.7 \times 10^7$$

$$K_b = \frac{[NH_4^+][OH^-]}{[NH_3]} = 1.75 \times 10^{-5}$$

$$K_w = [H_3O^+][OH^-] = 1.00 \times 10^{-14}$$

Three additional equations are needed. The first of these equations is a mass balance for NH_3.

$$C_{NH_3} = [NH_3] + [NH_4^+] + 2 \times [Ag(NH_3)_2^+]$$

Note that in writing this mass balance equation, the concentration of $Ag(NH_3)_2^+$ must be multiplied by 2 since two moles of NH_3 occurs per mole of $Ag(NH_3)_2^+$. The second additional equation is a mass balance on iodide and silver. Since AgI is the only source of I^- and Ag^+, every iodide in solution must have an associated silver ion; thus

$$[I^-] = [Ag^+] + [Ag(NH_3)_2^+]$$

Finally, the last equation is a charge balance equation

$$[Ag^+] + [Ag(NH_3)_2^+] + [NH_4^+] + [H_3O^+] = [I^-] + [OH^-]$$

Our problem looks challenging, but several assumptions greatly simplify the algebra. First, since the formation of the $Ag(NH_3)_2^+$ complex is favorable, we will assume that

$$[Ag^+] << [Ag(NH_3)_2^+]$$

Second, since NH_3 is a base, we will assume that

$$[H_3O^+] << [OH^-]$$

$$[NH_4^+] << [NH_3] + [Ag(NH_3)_2^+]$$

Finally, since K_{sp} is significantly smaller than β_2, it seems likely that the solubility of AgI is small and

$$[Ag(NH_3)_2^+] << [NH_3]$$

Using these assumptions allows us to simplify several equations. The mass balance for NH_3 is now

$$C_{NH_3} = [NH_3]$$

and the mass balance for I^- is

$$[I^-] = [Ag(NH_3)_2^+]$$

Simplifying the charge balance expression by dropping $[H_3O^+]$ and $[Ag^+]$, and replacing $[Ag(NH_3)_2^+]$ with $[I^-]$ gives

$$[NH_4^+] = [OH^-]$$

We continue by multiplying together the equations for K_{sp} and β_2, giving

$$K_{sp}\beta_2 = \frac{[Ag(NH_3)_2^+][I^-]}{[NH_3]^2} = 1.4 \times 10^{-9}$$

Substituting in the new mass balance equations for NH_3 and I^-

$$\frac{[I^-]^2}{(C_{NH_3})^2} = 1.4 \times 10^{-9}$$

and solving for the $[I^-]$ gives

$$\frac{[I^-]^2}{(0.10\,M)^2} = 1.4 \times 10^{-9}$$

$$[I^-] = 3.7 \times 10^{-6}\,M$$

Before accepting this answer, we first check our assumptions. Using the K_{sp} equation we calculate the $[Ag^+]$ to be

$$[Ag^+] = \frac{K_{sp}}{[I^-]} = \frac{8.3 \times 10^{-17}}{3.7 \times 10^{-6}} = 2.2 \times 10^{-11}\,M$$

From the simplified mass balance equation for I^-, we have

$$[Ag(NH_3)_2^+] = [I^-] = 3.7 \times 10^{-6}\,M$$

Our first assumption that the $[Ag^+]$ is significantly smaller than the $[Ag(NH_3)_2^+]$, therefore, is reasonable. Furthermore, our third assumption that the $[Ag(NH_3)_2^+]$ is significantly less than the $[NH_3]$ also is reasonable. Our second assumption was

$$[NH_4^+] << [NH_3] + [Ag(NH_3)_2^+]$$

To verify this assumption, we solve the K_b equation for $[NH_4^+]$

$$\frac{[NH_4^+][OH^-]}{[NH_3]} = \frac{[NH_4^+]^2}{0.10\,M} = 1.75 \times 10^{-5}$$

giving

$$[NH_4^+] = 1.3 \times 10^{-3}\,M$$

Although the $[NH_4^+]$ is not significantly smaller than the combined concentrations of NH_3 and $Ag(NH_3)_2^+$, the error is only about 1%. Since this is not an excessively large error, we will accept this approximation as reasonable.

Since one mole of AgI produces one mole of I^-, the solubility of AgI is the same as the concentration of iodide, or 3.7×10^{-6} mol/L.

6H Buffer Solutions

Adding as little as 0.1 mL of concentrated HCl to a liter of H_2O shifts the pH from 7.0 to 3.0. The same addition of HCl to a liter solution that is 0.1 M in both a weak acid and its conjugate weak base, however, results in only a negligible change in pH. Such solutions are called **buffers,** and their buffering action is a consequence of the relationship between pH and the relative concentrations of the conjugate weak acid/weak base pair.

buffer
A solution containing a conjugate weak acid/weak base pair that is resistant to a change in pH when a strong acid or strong base is added.

A mixture of acetic acid and sodium acetate is one example of an acid/base buffer. The equilibrium position of the buffer is governed by the reaction

$$CH_3COOH(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + CH_3COO^-(aq)$$

and its acid dissociation constant

$$K_a = \frac{[H_3O^+][CH_3COO^-]}{[CH_3COOH]} = 1.75 \times 10^{-5} \qquad \textbf{6.43}$$

The relationship between the pH of an acid–base buffer and the relative amounts of CH_3COOH and CH_3COO^- is derived by taking the negative log of both sides of equation 6.43 and solving for the pH

$$pH = pK_a + \log \frac{[CH_3COO^-]}{[CH_3COOH]} = 4.76 + \log \frac{[CH_3COO^-]}{[CH_3COOH]} \qquad \textbf{6.44}$$

Buffering occurs because of the logarithmic relationship between pH and the ratio of the weak base and weak acid concentrations. For example, if the equilibrium concentrations of CH_3COOH and CH_3COO^- are equal, the pH of the buffer is 4.76. If sufficient strong acid is added such that 10% of the acetate ion is converted to acetic acid, the concentration ratio $[CH_3COO^-]/[CH_3COOH]$ changes to 0.818, and the pH decreases to 4.67.

6H.1 Systematic Solution to Buffer Problems

Equation 6.44 is written in terms of the concentrations of CH_3COOH and CH_3COO^- at equilibrium. A more useful relationship relates the buffer's pH to the initial concentrations of weak acid and weak base. A general buffer equation can be derived by considering the following reactions for a weak acid, HA, and the salt of its conjugate weak base, NaA.

$$NaA(aq) \rightarrow Na^+(aq) + A^-(aq)$$

$$HA(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + A^-(aq)$$

$$2H_2O(\ell) \rightleftharpoons H_3O^+(aq) + OH^-(aq)$$

Since the concentrations of Na^+, A^-, HA, H_3O^+, and OH^- are unknown, five equations are needed to uniquely define the solution's composition. Two of these equations are given by the equilibrium constant expressions

$$K_a = \frac{[H_3O^+][A^-]}{[HA]}$$

$$K_w = [H_3O^+][OH^-]$$

The remaining three equations are given by mass balance equations on HA and Na^+

$$C_{HA} + C_{NaA} = [HA] + [A^-] \qquad \textbf{6.45}$$

$$C_{NaA} = [Na^+] \qquad \textbf{6.46}$$

and a charge balance equation

$$[H_3O^+] + [Na^+] = [OH^-] + [A^-]$$

Substituting equation 6.46 into the charge balance equation and solving for $[A^-]$ gives

$$[A^-] = C_{NaA} - [OH^-] + [H_3O^+] \qquad \textbf{6.47}$$

which is substituted into equation 6.45 to give the concentration of HA

$$[HA] = C_{HA} + [OH^-] - [H_3O^+] \qquad \textbf{6.48}$$

Finally, substituting equations 6.47 and 6.48 into the K_a equation for HA and solving for pH gives the general buffer equation

$$pH = pK_a + \log \frac{C_{NaA} - [OH^-] + [H_3O^+]}{C_{HA} + [OH^-] - [H_3O^+]}$$

If the initial concentrations of weak acid and weak base are greater than $[H_3O^+]$ and $[OH^-]$, the general equation simplifies to the **Henderson–Hasselbalch equation.**

Henderson–Hasselbalch equation
Equation showing the relationship between a buffer's pH and the relative amounts of the buffer's conjugate weak acid and weak base.

$$pH = pK_a + \log \frac{C_{NaA}}{C_{HA}} \qquad \textbf{6.49}$$

As in Example 6.13, the Henderson–Hasselbalch equation provides a simple way to calculate the pH of a buffer and to determine the change in pH upon adding a strong acid or strong base.

Example

EXAMPLE 6.13

Calculate the pH of a buffer that is 0.020 M in NH_3 and 0.030 M in NH_4Cl. What is the pH after adding 1.00 mL of 0.10 M NaOH to 0.10 L of this buffer?

SOLUTION

The acid dissociation constant for NH_4^+ is 5.70×10^{-10}; thus the initial pH of the buffer is

$$pH = 9.24 + \log \frac{C_{NH_3}}{C_{NH_4^+}} = 9.24 + \log \frac{0.020}{0.030} = 9.06$$

Adding NaOH converts a portion of the NH_4^+ to NH_3 due to the following reaction

$$NH_4^+(aq) + OH^-(aq) \rightleftharpoons NH_3(aq) + H_2O(\ell)$$

Since the equilibrium constant for this reaction is large, we may treat the reaction as if it went to completion. The new concentrations of NH_4^+ and NH_3 are therefore

$$C_{NH_4^+} = \frac{\text{moles } NH_4^+ - \text{moles } OH^-}{V_{tot}}$$

$$= \frac{(0.030\,\text{M})(0.10\,\text{L}) - (0.10\,\text{M})(1.00 \times 10^{-3}\,\text{L})}{0.101\,\text{L}} = 0.029\,\text{M}$$

$$C_{NH_3} = \frac{\text{moles } NH_3 + \text{moles } OH^-}{V_{tot}}$$

$$= \frac{(0.020\,\text{M})(0.10\,\text{L}) + (0.10\,\text{M})(1.00 \times 10^{-3}\,\text{L})}{0.101\,\text{L}} = 0.021\,\text{M}$$

Substituting the new concentrations into the Henderson–Hasselbalch equation gives a pH of

$$\text{pH} = 9.24 + \log \frac{0.021}{0.029} = 9.10$$

Multiprotic weak acids can be used to prepare buffers at as many different pH's as there are acidic protons. For example, a diprotic weak acid can be used to prepare buffers at two pH's and a triprotic weak acid can be used to prepare three different buffers. The Henderson–Hasselbalch equation applies in each case. Thus, buffers of malonic acid ($pK_{a1} = 2.85$ and $pK_{a2} = 5.70$) can be prepared for which

$$\text{pH} = 2.85 + \log \frac{C_{HM^-}}{C_{H_2M}}$$

$$\text{pH} = 5.70 + \log \frac{C_{M^{2-}}}{C_{HM^-}}$$

where H_2M, HM^-, and M^{2-} are the different forms of malonic acid.

The capacity of a buffer to resist a change in pH is a function of the absolute concentration of the weak acid and the weak base, as well as their relative proportions. The importance of the weak acid's concentration and the weak base's concentration is obvious. The more moles of weak acid and weak base that a buffer has, the more strong base or strong acid it can neutralize without significantly changing the buffer's pH. The relative proportions of weak acid and weak base affect the magnitude of the change in pH when adding a strong acid or strong base. Buffers that are equimolar in weak acid and weak base require a greater amount of strong acid or strong base to effect a change in pH of one unit.[2] Consequently, buffers are most effective to the addition of either acid or base at pH values near the pK_a of the weak acid.

Buffer solutions are often prepared using standard "recipes" found in the chemical literature.[3] In addition, computer programs have been developed to aid in the preparation of other buffers.[4] Perhaps the simplest means of preparing a buffer, however, is to prepare a solution containing an appropriate conjugate weak acid and weak base and measure its pH. The pH is easily adjusted to the desired pH by adding small portions of either a strong acid or a strong base.

Although this treatment of buffers was based on acid–base chemistry, the idea of a buffer is general and can be extended to equilibria involving complexation or redox reactions. For example, the Nernst equation for a solution containing Fe^{2+} and Fe^{3+} is similar in form to the Henderson–Hasselbalch equation.

$$E = E^\circ_{Fe^{3+}/Fe^{2+}} - 0.05916 \log \frac{[Fe^{2+}]}{[Fe^{3+}]}$$

Consequently, solutions of Fe^{2+} and Fe^{3+} are buffered to a potential near the standard-state reduction potential for Fe^{3+}.

6H.2 Representing Buffer Solutions with Ladder Diagrams

Ladder diagrams provide a simple graphical description of a solution's predominate species as a function of solution conditions. They also provide a convenient way to show the range of solution conditions over which a buffer is most effective. For ex-

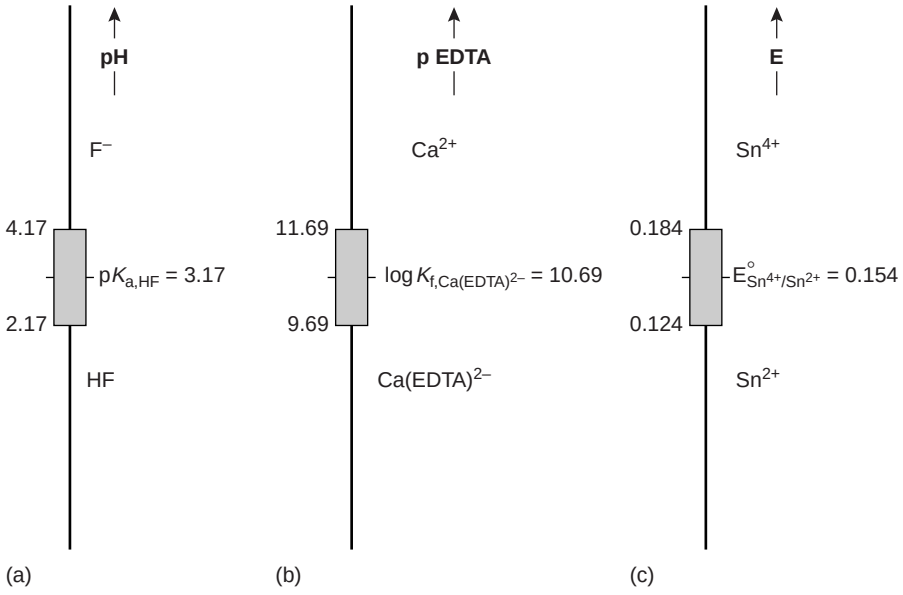


Figure 6.13
Ladder diagrams showing buffer regions for (a) HF/F^- acid–base buffer; (b) $Ca^{2+}/Ca(EDTA)^{2-}$ metal–ligand complexation buffer; and (c) SN^{4+}/Sn^{2+} oxidation–reduction buffer.

ample, an acid–base buffer can only exist when the relative abundance of the weak acid and its conjugate weak base are similar. For convenience, we will assume that an acid–base buffer exists when the concentration ratio of weak base to weak acid is between 0.1 and 10. Applying the Henderson–Hasselbalch equation

$$\text{pH} = \text{p}K_\text{a} + \log\frac{1}{10} = \text{p}K_\text{a} - 1$$

$$\text{pH} = \text{p}K_\text{a} + \log\frac{10}{1} = \text{p}K_\text{a} + 1$$

shows that acid–base buffer exists within the range of $\text{pH} = \text{p}K_\text{a} \pm 1$. In the same manner, it is easy to show that a complexation buffer for the metal–ligand complex ML_n exists when $\text{pL} = \log K_\text{f} \pm 1$, and that a redox buffer exists for $E = E^\circ \pm (0.05916/n)$. Ladder diagrams showing buffer regions for several equilibria are shown in Figure 6.13.

6I Activity Effects

Suppose you need to prepare a buffer with a pH of 9.36. Using the Henderson–Hasselbalch equation, you calculate the amounts of acetic acid and sodium acetate needed and prepare the buffer. When you measure the pH, however, you find that it is 9.25. If you have been careful in your calculations and measurements, what can account for the difference between the obtained and expected pHs? In this section, we will examine an important limitation to our use of equilibrium constants and learn how this limitation can be corrected.

Careful measurements of the solubility of $AgIO_3$ show that it increases in the presence of KNO_3, even though neither K^+ or NO_3^- participates in the solubility reaction.[5] Clearly the equilibrium position for the reaction

$$AgIO_3(s) \rightleftharpoons Ag^+(aq) + IO_3^-(aq)$$

depends on the composition of the solution. When the solubility product for $AgIO_3$ is calculated using the equilibrium concentrations of Ag^+ and IO_3^-

$$K_{sp} = [Ag^+][IO_3^-]$$

its apparent value increases when an inert electrolyte such as KNO_3 is added.

Why should adding an inert electrolyte affect the equilibrium position of a chemical reaction? We can explain the effect of KNO_3 on the solubility of $AgIO_3$ by considering the reaction on a microscopic scale. The solution in which equilibrium is established contains a variety of cations and anions—K^+, Ag^+, H_3O^+, NO_3^-, IO_3^- and OH^-. Although the solution is homogeneous, on the average, there are more anions in regions near Ag^+ ions, and more cations in regions near IO_3^- ions. Thus, Ag^+ and IO_3^- are surrounded by charged ionic atmospheres that partially screen the ions from each other. The formation of $AgIO_3$ requires the disruption of the ionic atmospheres surrounding the Ag^+ and IO_3^- ions. Increasing the concentrations of ions in solution, by adding KNO_3, increases the size of these ionic atmospheres. Since more energy is now required to disrupt the ionic atmospheres, there is a decrease in the formation of $AgIO_3$, and an apparent increase in the equilibrium constant.

ionic strength
A quantitative method for reporting the ionic composition of a solution that takes into account the greater effect of more highly charged ions (μ).

The ionic composition of a solution frequently is expressed by its **ionic strength,** μ

$$\mu = \frac{1}{2}\sum_i c_i z_i^2$$

where c_i and z_i are the concentration and charge of the ith ion.

Example

EXAMPLE 6.14

Calculate the ionic strength of 0.10 M NaCl. Repeat the calculation for a solution of 0.10 M Na_2SO_4.

SOLUTION

The ionic strength for 0.10 M NaCl is

$$\mu = \frac{1}{2}([Na^+](+1)^2 + [Cl^-](-1)^2) = \frac{1}{2}[(0.10)(+1)^2 + (0.10)(-1)^2] = 0.10\,M$$

For 0.10 M Na_2SO_4, the ionic strength is

$$\mu = \frac{1}{2}([Na^+](+1)^2 + [SO_4^{2-}](-2)^2) = \frac{1}{2}[(0.20)(+1)^2 + (0.10)(-2)^2] = 0.30\,M$$

Note that the unit for ionic strength is molarity, but that the molar ionic strength need not match the molar concentration of the electrolyte. For a 1:1 electrolyte, such as NaCl, ionic strength and molar concentration are identical. The ionic strength of a 2:1 electrolyte, such as Na_2SO_4, is three times larger than the electrolyte's molar concentration.

activity
True thermodynamic constants use a species activity in place of its molar concentration (a).

The true thermodynamic equilibrium constant is a function of **activity** rather than concentration. The activity of a species, a_A, is defined as the product of its molar concentration, [A], and a solution-dependent **activity coefficient,** γ_A.

activity coefficient
The number that when multiplied by a species' concentration gives that species' activity (γ).

$$a_A = [A]\gamma_A$$

The true thermodynamic equilibrium constant, K_{sp}, for the solubility of $AgIO_3$, therefore, is

$$K_{sp} = (a_{Ag}+)(a_{IO_3^-}) = [Ag^+][IO_3^-](\gamma_{Ag}+)(\gamma_{IO_3^-})$$

To accurately calculate the solubility of $AgIO_3$, we must know the activity coefficients for Ag^+ and IO_3^-.

For gases, pure solids, pure liquids, and nonionic solutes, activity coefficients are approximately unity under most reasonable experimental conditions. For reactions involving only these species, differences between activity and concentration are negligible. Activity coefficients for ionic solutes, however, depend on the ionic composition of the solution. It is possible, using the extended Debye–Hückel theory,* to calculate activity coefficients using equation 6.50

$$-\log \gamma_A = \frac{0.51 \times z_A^2 \times \sqrt{\mu}}{1 + 3.3 \times \alpha_A \times \sqrt{\mu}} \qquad \textbf{6.50}$$

where Z_A is the charge of the ion, α_A is the effective diameter of the hydrated ion in nanometers (Table 6.1), μ is the solution's ionic strength, and 0.51 and 3.3 are constants appropriate for aqueous solutions at 25 °C.

Several features of equation 6.50 deserve mention. First, as the ionic strength approaches zero, the activity coefficient approaches a value of one. Thus, in a solution where the ionic strength is zero, an ion's activity and concentration are identical. We can take advantage of this fact to determine a reaction's thermodynamic equilibrium constant. The equilibrium constant based on concentrations is measured for several increasingly smaller ionic strengths and the results extrapolated

Table 6.1 Effective Diameters (α) for Selected Inorganic Cations and Anions

Ion	Effective Diameter (nm)
H_3O^+	0.9
Li^+	0.6
Na^+, IO_3^-, HSO_3^-, HCO_3^-, $H_2PO_4^-$	0.45
OH^-, F^-, SCN^-, HS^-, ClO_3^-, ClO_4^-, MnO_4^-	0.35
K^+, Cl^-, Br^-, I^-, CN^-, NO_2^-, NO_3^-	0.3
Cs^+, Tl^+, Ag^+, NH_4^+	0.25
Mg^{2+}, Be^{2+}	0.8
Ca^{2+}, Cu^{2+}, Zn^{2+}, Sn^{2+}, Mn^{2+}, Fe^{2+}, Ni^{2+}, Co^{2+}	0.6
Sr^{2+}, Ba^{2+}, Cd^{2+}, Hg^{2+}, S^{2-}	0.5
Pb^{2+}, CO_3^{2-}, SO_3^{2-}	0.45
Hg_2^{2+}, SO_4^{2-}, $S_2O_3^{2-}$, CrO_4^{2-}, HPO_4^{2-}	0.40
Al^{3+}, Fe^{3+}, Cr^{3+}	0.9
PO_4^{3-}, $Fe(CN)_6^{3-}$	0.4
Zr^{4+}, Ce^{4+}, Sn^{4+}	1.1
$Fe(CN)_6^{4-}$	0.5

Source: Values from Kielland, J. *J. Am. Chem. Soc.* **1937,** *59,* 1675.

*See any standard textbook on physical chemistry for more information on the Debye–Hückel theory and its application to solution equilibrium

back to zero ionic strength to give the thermodynamic equilibrium constant. Second, activity coefficients are smaller, and thus activity effects are more important, for ions with higher charges and smaller effective diameters. Finally, the extended Debye–Hückel equation provides reasonable activity coefficients for ionic strengths of less than 0.1. Modifications to the extended Debye–Hückel equation, which extend the calculation of activity coefficients to higher ionic strength, have been proposed.[6]

Example

EXAMPLE 6.15

Calculate the solubility of $Pb(IO_3)_2$ in a matrix of 0.020 M $Mg(NO_3)_2$.

SOLUTION

We begin by calculating the ionic strength of the solution. Since $Pb(IO_3)_2$ is only sparingly soluble, we will assume that its contribution to the ionic strength can be ignored; thus

$$\mu = \frac{1}{2}[(0.20\,\text{M})(+2)^2 + (0.040\,\text{M})(-1)^2] = 0.060\,\text{M}$$

Activity coefficients for Pb^{2+} and I^- are calculated using equation 6.50

$$-\log \gamma_{Pb^{2+}} = \frac{0.51 \times (+2)^2 \times \sqrt{0.060}}{1 + 3.3 \times 0.45 \times \sqrt{0.060}} = 0.366$$

giving an activity coefficient for Pb^{2+} of 0.43. A similar calculation for IO_3^- gives its activity coefficient as 0.81. The equilibrium constant expression for the solubility of PbI_2 is

$$K_{sp} = [Pb^{2+}][IO_3^-]^2 \gamma_{Pb^{2+}} \gamma_{IO_3^-} = 2.5 \times 10^{-13}$$

Letting

$$[Pb^{2+}] = x \qquad \text{and} \qquad [IO_3^-] = 2x$$

we have

$$(x)(2x)^2(0.45)(0.81)^2 = 2.5 \times 10^{-13}$$

Solving for x gives a value of 6.0×10^{-5} or a solubility of 6.0×10^{-5} mol/L. This compares to a value of 4.0×10^{-5} mol/L when activity is ignored. Failing to correct for activity effects underestimates the solubility of PbI_2 in this case by 33%.

As this example shows, failing to correct for the effect of ionic strength can lead to significant differences between calculated and actual concentrations. Nevertheless, it is not unusual to ignore activities and assume that the equilibrium constant is expressed in terms of concentrations. There is a practical reason for this—in an analysis one rarely knows the composition, much less the ionic strength of a sample solution. Equilibrium calculations are often used as a guide when developing an analytical method. Only by conducting the analysis and evaluating the results can we judge whether our theory matches reality.

Colorplate 3 provides a visual demonstration of the effect of ionic strength on the equilibrium reaction $Fe^{3+}(aq) + SCN^-(aq) \rightleftharpoons Fe(SCN)^{2+}(aq)$

6J Two Final Thoughts About Equilibrium Chemistry

In this chapter we have reviewed and extended our understanding of equilibrium chemistry. We also have developed several tools for evaluating the composition of a system at equilibrium. These tools differ in how accurately they allow us to answer questions involving equilibrium chemistry. They also differ in their ease of use. An important part of having several tools available to you is knowing when to use them. If you need to know whether a reaction is favorable, or the approximate pH of a solution, a ladder diagram may be sufficient to meet your needs. On the other hand, if you require an accurate estimate of a compound's solubility, a rigorous calculation using the systematic approach and activity coefficients is necessary.

Finally, a consideration of equilibrium chemistry can only help us decide what reactions are favorable. Knowing that a reaction is favorable does not guarantee that the reaction will occur. How fast a reaction approaches its equilibrium position does not depend on the magnitude of the equilibrium constant. The rate of a chemical reaction is a kinetic, not a thermodynamic, phenomenon. Kinetic effects and their application in analytical chemistry are discussed in Chapter 13.

6K KEY TERMS

acid (*p. 140*)
acid dissociation constant (*p. 140*)
activity (*p. 172*)
activity coefficient (*p. 172*)
amphiprotic (*p. 142*)
base (*p. 140*)
base dissociation constant (*p. 141*)
buffer (*p. 167*)
charge balance equation (*p. 159*)
common ion effect (*p. 158*)
cumulative formation constant (*p. 144*)
dissociation constant (*p. 144*)
enthalpy (*p. 137*)
entropy (*p. 137*)
equilibrium (*p. 136*)
equilibrium constant (*p. 138*)
formation constant (*p. 144*)
Gibb's free energy (*p. 137*)
Henderson–Hasselbalch equation (*p. 169*)
ionic strength (*p. 172*)
ladder diagram (*p. 150*)
Le Châtelier's principle (*p. 148*)
ligand (*p. 144*)
mass balance equation (*p. 159*)
Nernst equation (*p. 146*)
oxidation (*p. 146*)
oxidizing agent (*p. 146*)
pH (*p. 142*)
precipitate (*p. 139*)
redox reaction (*p. 145*)
reducing agent (*p. 146*)
reduction (*p. 146*)
solubility product (*p. 140*)
standard state (*p. 137*)
stepwise formation constant (*p. 144*)

6L SUMMARY

Analytical chemistry is more than a collection of techniques; it is the application of chemistry to the analysis of samples. As you will see in later chapters, almost all analytical methods use chemical reactivity to accomplish one or more of the following—dissolve the sample, separate analytes and interferents, transform the analyte to a more useful form, or provide a signal. Equilibrium chemistry and thermodynamics provide us with a means for predicting which reactions are likely to be favorable.

The most important types of reactions are precipitation reactions, acid–base reactions, metal–ligand complexation reactions, and redox reactions. In a precipitation reaction two or more soluble species combine to produce an insoluble product called a precipitate. The equilibrium properties of a precipitation reaction are described by a solubility product.

Acid–base reactions occur when an acid donates a proton to a base. The equilibrium position of an acid–base reaction is described using either the dissociation constant for the acid, K_a, or the dissociation constant for the base, K_b. The product of K_a and K_b for an acid and its conjugate base is K_w (water's dissociation constant).

Ligands have electron pairs that they can donate to a metal ion, forming a metal–ligand complex. The formation of the metal–ligand complex ML_2, for example, may be described by a stepwise formation constant in which each ligand is added one at a time; thus, K_1 represents the addition of the first ligand to M, and K_2 represents the addition of the second ligand to ML. Alternatively, the formation of ML_2 can be described by a cumulative, or overall formation constant, β_2, in which both ligands are added to M.

In a redox reaction, one of the reactants is oxidized while another reactant is reduced. Equilibrium constants are rarely used when characterizing redox reactions. Instead, we use the electrochemical potential, positive values of which indicate a favorable reaction. The Nernst equation relates this potential to the concentrations of reactants and products.

Le Châtelier's principle provides a means for predicting how systems at equilibrium respond to a change in conditions. When a stress is applied to an equilibrium by adding a reactant or product, by adding a reagent that reacts with one of the reactants or products, or by changing the volume, the system responds by moving in the direction that relieves the stress.

You should be able to describe a system at equilibrium both qualitatively and quantitatively. Rigorous solutions to equilibrium problems can be developed by combining equilibrium constant expressions with appropriate mass balance and charge balance equations. Using this systematic approach, you can solve some quite complicated equilibrium problems. When a less rigorous answer is needed, a ladder diagram may help you decide the equilibrium system's composition.

Solutions containing a weak acid and its conjugate base show only a small change in pH upon the addition of small amounts of strong acid or strong base. Such solutions are called buffers. Buffers can also be formed using a metal and its metal–ligand complex, or an oxidizing agent and its conjugate reducing agent. Both the systematic approach to solving equilibrium problems and ladder diagrams can be used to characterize a buffer.

A quantitative solution to an equilibrium problem may give an answer that does not agree with the value measured experimentally. This result occurs when the equilibrium constant based on concentrations is matrix-dependent. The true, thermodynamic equilibrium constant is based on the activities, a, of the reactants and products. A species' activity is related to its molar concentration by an activity coefficient, γ, where $a_i = \gamma_i[\ ]_i$. Activity coefficients often can be calculated, making possible a more rigorous treatment of equilibria.

6M *Suggested* EXPERIMENTS

Experiments

The following experiments involve the experimental determination of equilibrium constants and, in some cases, demonstrate the importance of activity effects.

"The Effect of Ionic Strength on an Equilibrium Constant (A Class Study)." In J. A. Bell, ed. *Chemical Principles in Practice.* Addison-Wesley: Reading, MA, 1967.

In this experiment the equilibrium constant for the dissociation of bromocresol green is measured at several ionic strengths. Results are extrapolated to zero ionic strength to find the thermodynamic equilibrium constant.

"Equilibrium Constants for Calcium Iodate Solubility and Iodic Acid Dissociation." In J. A. Bell, ed. *Chemical Principles in Practice.* Addison-Wesley: Reading, MA, 1967.

The effect of pH on the solubility of $Ca(IO_3)_2$ is studied in this experiment.

"The Solubility of Silver Acetate." In J. A. Bell, ed. *Chemical Principles in Practice.* Addison-Wesley: Reading, MA, 1967.

In this experiment the importance of the soluble silver acetate complexes $AgCH_3COO(aq)$ and $Ag(CH_3COO)_2^-(aq)$ in describing the solubility of $AgCH_3COO(s)$ is investigated.

Green, D. B.; Rechtsteiner, G.; Honodel, A. "Determination of the Thermodynamic Solubility Product, K_{sp}, of PbI_2 Assuming Nonideal Behavior," *J. Chem. Educ.* **1996,** *73,* 789–792.

The thermodynamic solubility product for PbI_2 is determined in this experiment by measuring its solubility at several ionic strengths.

6N PROBLEMS

1. Write equilibrium constant expressions for the following reactions. Determine the value for the equilibrium constant for each reaction using appropriate equilibrium constants from Appendix 3.

a. $NH_3(aq) + HCl(aq) \rightleftharpoons NH_4^+(aq) + Cl^-(aq)$

b. $PbI_2(s) + S^{2-}(aq) \rightleftharpoons PbS(s) + 2I^-(aq)$

c. $CdY^{2-}(aq) + 4CN^-(aq) \rightleftharpoons Cd(CN)_4^{2-}(aq) + Y^{4-}(aq)$ [Y^{4-} is EDTA]

d. $AgCl(s) + 2NH_3(aq) \rightleftharpoons Ag(NH_3)_2^+(aq) + Cl^-(aq)$

e. $BaCO_3(s) + 2H_3O^+(aq) \rightleftharpoons Ba^{2+}(aq) + H_2CO_3(aq) + 2H_2O(\ell)$

2. Using a ladder diagram, explain why the following reaction

$$H_3PO_4(aq) + F^-(aq) \rightleftharpoons HF(aq) + H_2PO_4^-(aq)$$

is favorable, whereas

$$H_3PO_4(aq) + 2F^-(aq) \rightleftharpoons 2HF(aq) + H_2PO_4^{2-}(aq)$$

is unfavorable. Determine the equilibrium constant for these reactions, and verify that they are consistent with your ladder diagram.

3. Calculate the potential for the following redox reaction when the $[Fe^{3+}] = 0.050$ M, $[Fe^{2+}] = 0.030$ M, $[Sn^{2+}] = 0.015$ M and $[Sn^{4+}] = 0.020$ M

$$2Fe^{3+}(aq) + Sn^{2+}(aq) \rightleftharpoons Sn^{4+}(aq) + 2Fe^{2+}(aq)$$

4. Balance the following redox reactions, and calculate the standard-state potential and the equilibrium constant for each. Assume that the $[H_3O^+]$ is 1 M for acidic solutions, and that the $[OH^-]$ is 1 M for basic solutions.
   a. $MnO_4^-(aq) + H_2SO_3(aq) \rightleftharpoons Mn^{2+}(aq) + SO_4^{2-}(aq)$ (acidic solution)
   b. $IO_3^-(aq) + I^-(aq) \rightleftharpoons I_2(s)$ (acidic solution)
   c. $ClO^-(aq) + I^- \rightleftharpoons IO_3^-(aq) + Cl^-(aq)$ (basic solution)

5. Sulfur can be determined quantitatively by oxidizing to SO_4^{2-} and precipitating as $BaSO_4$. The solubility reaction for $BaSO_4$ is

$$BaSO_4(s) \rightleftharpoons Ba^{2+}(aq) + SO_4^{2-}(aq)$$

How will the solubility of $BaSO_4$ be affected by (a) decreasing the pH of the solution; (b) adding $BaCl_2$; (c) decreasing the volume of the solution?

6. Write charge balance and mass balance equations for the following solutions
   a. 0.1 M NaCl
   b. 0.1 M HCl
   c. 0.1 M HF
   d. 0.1 M NaH_2PO_4
   e. $MgCO_3$ (saturated solution)
   f. 0.1 M $Ag(CN)_2^-$ (from $AgNO_3$ and KCN)
   g. 0.1 M HCl and 0.050 M $NaNO_2$

7. Using the systematic approach, calculate the pH of the following solutions
   a. 0.050 M $HClO_4$
   b. 1.00×10^{-7} M HCl
   c. 0.025 M HClO
   d. 0.010 M HCOOH
   e. 0.050 M $Ba(OH)_2$
   f. 0.010 M C_5H_5N

8. Construct ladder diagrams for the following diprotic weak acids (H_2L), and estimate the pH of 0.10 M solutions of H_2L, HL^-, and L^{2-}. Using the systematic approach, calculate the pH of each of these solutions.
   a. maleic acid
   b. malonic acid
   c. succinic acid

9. Ignoring activity effects, calculate the solubility of Hg_2Cl_2 in the following
   a. A saturated solution of Hg_2Cl_2
   b. 0.025 M $Hg_2(NO_3)_2$ saturated with Hg_2Cl_2
   c. 0.050 M NaCl saturated with Hg_2Cl_2

10. The solubility of CaF_2 is controlled by the following two reactions

$$CaF_2(s) \rightleftharpoons Ca^{2+}(aq) + F^-(aq)$$

$$HF(aq) + H_2O(\ell) \rightleftharpoons H_3O^+(aq) + F^-(aq)$$

Calculate the solubility of CaF_2 in a solution buffered to a pH of 7.00. Use a ladder diagram to help simplify the calculations. How would your approach to this problem change if the pH is buffered to 2.00? What is the solubility of CaF_2 at this pH?

11. Calculate the solubility of $Mg(OH)_2$ in a solution buffered to a pH of 7.00. How does this compare with its solubility in unbuffered water?

12. Calculate the solubility of Ag_3PO_4 in a solution buffered to a pH of 9.00.

13. Determine the equilibrium composition of saturated solution of AgCl. Assume that the solubility of AgCl is influenced by the following reactions.

$$AgCl(s) \rightleftharpoons Ag^+(aq) + Cl^-(aq)$$

$$Ag^+(aq) + Cl^-(aq) \rightleftharpoons AgCl(aq)$$

$$AgCl(aq) + Cl^-(aq) \rightleftharpoons AgCl_2^-(aq)$$

14. Calculate the ionic strength of the following solutions
   a. 0.050 M NaCl
   b. 0.025 M $CuCl_2$
   c. 0.10 M Na_2SO_4

15. Repeat the calculations in problem 9, this time correcting for activity effects.

16. With the permission of your instructor, carry out the following experiment. In a beaker, mix equal volumes of 0.001 M NH_4SCN and 0.001 M $FeCl_3$ (the latter solution must be acidified with concentrated HNO_3 at a ratio of 4 drops/L to prevent the precipitation of $Fe(OH)_3$). Divide solution in half, and add solid KNO_3 to one portion at a ratio of 4 g per 100 mL. Compare the colors of the two solutions (see Color Plate 3), and explain why they are different. The relevant reaction is

$$Fe^{3+}(aq) + SCN^-(aq) \rightleftharpoons Fe(SCN)^{2+}(aq)$$

17. Over what pH range do you expect $Ca_3(PO_4)_2$ to have its minimum solubility?

18. Construct ladder diagrams for the following systems, and describe the information that can be obtained from each
   a. HF and H_3PO_4
   b. $Ag(CN)_2^-$, $Ni(CN)_4^{2-}$ and $Fe(CN)_6^{4-}$
   c. $Cr_2O_7^{2-}/Cr^{3+}$ and Fe^{3+}/Fe^{2+}

19. Calculate the pH of the following acid–base buffers
   a. 100 mL of 0.025 M formic acid and 0.015 M sodium formate
   b. 50.00 mL of 0.12 M NH_3 and 3.50 mL of 1.0 M HCl
   c. 5.00 g of Na_2CO_3 and 5.00 g of $NaHCO_3$ in 0.100 L

20. Calculate the pH of the buffers in problem 19 after adding 5.0×10^{-4} mol of HCl.

21. Calculate the pH of the buffers in problem 19 after adding 5.0×10^{-4} mol of NaOH.

22. Consider the following hypothetical complexation reaction between a metal, M, and a ligand, L

$$M(aq) + L(aq) \rightleftharpoons ML(aq)$$

with a formation constant of 1.5×10^8. Derive an equation, similar to the Henderson–Hasselbalch equation, which relates pM to the concentrations of L and ML. What will be the pM for a solution containing 0.010 mol of M and 0.020 mol of L? What will the pM be if 0.002 mol of M are added?

23. A redox buffer contains an oxidizing agent and its conjugate reducing agent. Calculate the potential of a solution containing 0.010 mol of Fe^{3+} and 0.015 mol of Fe^{2+}. What is the potential if sufficient oxidizing agent is added such that 0.002 mol of Fe^{2+} is converted to Fe^{3+}?

6O SUGGESTED READINGS

A lucid discussion of Berthollet's discovery of the reversibility of reactions is found in

Roots-Bernstein, R. S. *Discovering.* Harvard University Press: Cambridge, MA, 1989.

The following texts and articles provide additional coverage of equilibrium chemistry and the systematic approach to solving equilibrium problems.

Butler, J. N. *Ionic Equilibria: A Mathematical Approach.* Addison-Wesley: Reading, MA, 1964.

Butler, J. N. *Solubility and pH Calculations.* Addison-Wesley: Reading, MA, 1973.

Chaston, S. "Calculating Complex Equilibrium Concentrations by a Next Guess Factor Method," *J. Chem. Educ.* **1993,** *70,* 622–624.

Fernando, Q.; Ryan, M. D. *Calculations in Analytical Chemistry,* Harcourt Brace Jovanovich: New York, 1982.

Freiser, H. *Concepts and Calculations in Analytical Chemistry,* CRC Press: Boca Raton, 1992.

Freiser, H.; Fernando, Q. *Ionic Equilibria in Analytical Chemistry,* Wiley: New York, 1963.

Gordus, A. A. "Chemical Equilibrium I. The Thermodynamic Equilibrium Concept," *J. Chem. Educ.* **1991,** *68,* 138–140.

Gordus, A. A. "Chemical Equilibrum II. Deriving an Exact Equilibrium Equation," *J. Chem. Educ.* **1991,** *68,* 215–217.

Gordus, A. A. "Chemical Equilibrium III. A Few Math Tricks," *J. Chem. Educ.* **1991,** *68,* 291–293.

Gordus, A. A. "Chemical Equilibrium IV. Weak Acids and Bases," *J. Chem. Educ.* **1991,** *68,* 397–399.

Gordus, A. A. "Chemical Equilibrium VI. Buffer Solutions," *J. Chem. Educ.* **1991,** *68,* 656–658.

Gordus, A. A. "Chemical Equilibrium VII. Precipitates," *J. Chem. Educ.* **1991,** *68,* 927–930.

Gordus, A. A. *Schaum's Outline of Analytical Chemistry.* McGraw-Hill: New York, 1985.

Olivieri, A. C. "Solution of Acid–Base Equilibria by Successive Approximations," *J. Chem. Educ.* **1990,** *67,* 229–231.

Ramette, R. W. *Chemical Equilibrium and Analysis.* Addison-Wesley: Reading, MA, 1981.

Thomson, B. M.; Kessick, M. A. "On the Preparation of Buffer Solutions," *J. Chem. Educ.* **1981,** *58,* 743–746.

Weltin, E. "Are the Equilibrium Concentrations for a Chemical Reaction Always Uniquely Determined by the Initial Concentrations?" *J. Chem. Educ.* **1990,** *67,* 548.

Weltin, E. "A Numerical Method to Calculate Equilibrium Concentrations for Single-Equation Systems," *J. Chem. Educ.* **1991,** *68,* 486–487.

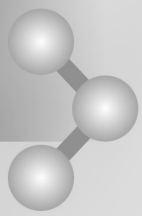
Weltin, E. "Calculating Equilibrium Concentrations," *J. Chem. Educ.* **1992,** *69,* 393–396.

Weltin, E. "Calculating Equilibrium Concentrations for Stepwise Binding of Ligands and Polyprotic Acid-Base Systems," *J. Chem. Educ.* **1993,** *70,* 568–571.

Weltin, E. "Equilibrium Calculations are Easier Than You Think— But You Do Have to Think!" *J. Chem. Educ.* **1993,** *70,* 571–573.

6P REFERENCES

1. (a) Runo, J. R.; Peters, D. G. *J. Chem. Educ.* **1993,** *70,* 708–713; (b) Vale, J.; Fernandez-Pereira, C.; Alcalde, M. *J. Chem. Educ.* **1993,** *70,* 790–795.
2. Van Slyke, D. D. *J. Biol. Chem.* **1922,** *52,* 525–570.
3. (a) Bower, V. E.; Bates, R. G. *J. Res. Natl. Bur. Stand. (U. S.)* **1955,** *55,* 197–200; (b) Bates, R. G. *Ann. N. Y. Acad. Sci.* **1961,** *92,* 341–356; (c) Bates, R. G. *Determination of pH,* 2nd ed. Wiley-Interscience: New York, **1973,** p. 73.
4. Lambert, W. J. *J. Chem. Educ.* **1990,** *67,* 150–153.
5. Kolthoff, I. M.; Lingane, J. *J. Phys. Chem.* **1938,** *42,* 133–140.
6. Davies, C. W. *Ion Association.* Butterworth: London, 1962.



Chapter 7

Obtaining and Preparing Samples for Analysis

When we first use an analytical method to solve a problem, it is not unusual to find that our results are of questionable accuracy or so imprecise as to be meaningless. Looking back we may find that nothing in the method seems amiss. In designing the method we considered sources of determinate and indeterminate error and took appropriate steps, such as including a reagent blank and calibrating our instruments, to minimize their effect. Why, then, might a carefully designed method give such poor results? One explanation is that we may not have accounted for errors associated with the sample. When we collect the wrong sample or lose analyte while preparing the sample for analysis, we introduce a determinate source of error. If we do not collect enough samples or collect samples of the wrong size, the precision of the analysis may suffer. In this chapter we consider how collecting samples and preparing them for analysis can affect the accuracy and precision of our results.

7A The Importance of Sampling

When a manufacturer produces a chemical they wish to list as ACS Reagent Grade, they must demonstrate that it conforms to specifications established by the American Chemical Society (ACS). For example, ACS specifications for $NaHCO_3$ require that the concentration of iron be less than or equal to 0.001% w/w. To verify that a production lot meets this standard, the manufacturer performs a quantitative analysis, reporting the result on the product's label. Because it is impractical to analyze the entire production lot, its properties are estimated from a limited sampling. Several samples are collected and analyzed, and the resulting mean, $\overline{X}$, and standard deviation, s, are used to establish a confidence interval for the production lot's true mean, μ

$$\mu = \overline{X} \pm \frac{ts}{\sqrt{n}} \qquad \textbf{7.1}$$

where n is the number of samples, and t is a statistical factor whose value is determined by the number of samples and the desired confidence level.*

Selecting a sample introduces a source of determinate error that cannot be corrected during the analysis. If a sample does not accurately represent the population from which it is drawn, then an analysis that is otherwise carefully conducted will yield inaccurate results. Sampling errors are introduced whenever we extrapolate from a sample to its target population. To minimize sampling errors we must collect the right sample.

Even when collecting the right sample, indeterminate or random errors in sampling may limit the usefulness of our results. Equation 7.1 shows that the width of a confidence interval is directly proportional to the standard deviation. The overall standard deviation for an analysis, s_o, is determined by random errors affecting each step of the analysis. For convenience, we divide the analysis into two steps. Random errors introduced when collecting samples are characterized by a standard deviation for sampling, s_s. The standard deviation for the analytical method, s_m, accounts for random errors introduced when executing the method's procedure. The relationship among s_o, s_s, and s_m is given by a propagation of random error

$$s_o^2 = s_m^2 + s_s^2 \qquad \textbf{7.2}$$

Equation 7.2 shows that an analysis' overall variance may be limited by either the analytical method or sample collection. Unfortunately, analysts often attempt to minimize overall variance by improving only the method's precision. This is futile, however, if the standard deviation for sampling is more than three times greater than that for the method.[1] Figure 7.1 shows how the ratio s_m/s_s affects the percentage of overall variance attributable to the method. When the method's standard deviation is one third of that for sampling, indeterminate method errors explain only 10% of the overall variance. Attempting to improve the analysis by decreasing s_m provides only a nominal change in the overall variance.

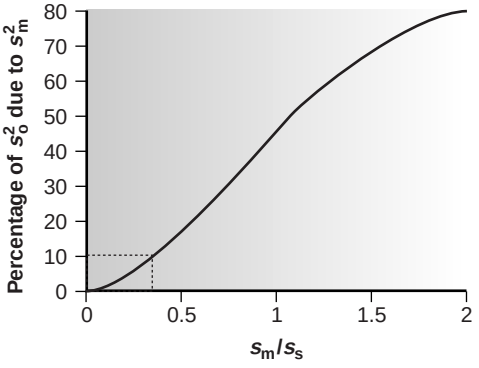


Figure 7.1
Percent of overall variance (s_o^2) due to the method as a function of the relative magnitudes of the standard deviation of the method and the standard deviation of sampling (s_m/s_s). The dotted lines show that the variance due to the method accounts for 10% of the overall variance when $s_s = 3 \times s_m$.

*Values for t can be found in Appendix 1B.

Example

EXAMPLE 7.1

A quantitative analysis for an analyte gives a mean concentration of 12.6 ppm. The standard deviation for the method is found to be 1.1 ppm, and that due to sampling is 2.1 ppm. (a) What is the overall variance for the analysis? (b) By how much does the overall variance change if s_m is improved by 10% to 0.99 ppm? (c) By how much does the overall variance change if s_s is improved by 10% to 1.9 ppm?

SOLUTION

(a) The overall variance is

$$s_o^2 = s_m^2 + s_s^2 = (1.1)^2 + (2.1)^2 = 1.21 + 4.41 = 5.62 \approx 5.6$$

(b) Improving the method's standard deviation changes the overall variance to

$$s_o^2 = (0.99)^2 + (2.1)^2 = 0.98 + 4.41 = 5.39 \approx 5.4$$

Thus, a 10% improvement in the method's standard deviation changes the overall variance by approximately 4%.

(c) Changing the standard deviation for sampling

$$s_o^2 = (1.1)^2 + (1.9)^2 = 1.21 + 3.61 = 4.82 \approx 4.8$$

improves the overall variance by almost 15%. As expected, since s_s is larger than s_m, a more significant improvement in the overall variance is realized when we focus our attention on sampling problems.

To determine which step has the greatest effect on the overall variance, both s_m^2 and s_s^2 must be known. The analysis of replicate samples can be used to estimate the overall variance. The variance due to the method is determined by analyzing a standard sample, for which we may assume a negligible sampling variance. The variance due to sampling is then determined by difference.

Example

EXAMPLE 7.2

The following data were collected as part of a study to determine the effect of sampling variance on the analysis of drug animal-feed formulations.[2]

% Drug (w/w)			% Drug (w/w)		
0.0114	0.0099	0.0105	0.0105	0.0109	0.0107
0.0102	0.0106	0.0087	0.0103	0.0103	0.0104
0.0100	0.0095	0.0098	0.0101	0.0101	0.0103
0.0105	0.0095	0.0097			

The data on the left were obtained under conditions in which random errors in sampling and the analytical method contribute to the overall variance. The data on the right were obtained in circumstances in which the sampling variance is known to be insignificant. Determine the overall variance and the contributions from sampling and the analytical method.

SOLUTION

The overall variance, s_o^2, is determined using the data on the left and is equal to 4.71×10^{-7}. The method's contribution to the overall variance, s_m^2, is determined using the data on the right and is equal to 7.00×10^{-8}. The variance due to sampling, s_s^2, is therefore

$$s_s^2 = s_o^2 - s_m^2 = 4.71 \times 10^{-7} - 7.00 \times 10^{-8} = 4.01 \times 10^{-7}$$

7B Designing A Sampling Plan

sampling plan
A plan that ensures that a representative sample is collected.

A **sampling plan** must support the goals of an analysis. In characterization studies a sample's purity is often the most important parameter. For example, a material scientist interested in the surface chemistry of a metal is more likely to select a freshly exposed surface, created by fracturing the sample under vacuum, than a surface that has been exposed to the atmosphere for an extended time. In a qualitative analysis the sample's composition does not need to be identical to that of the substance being analyzed, provided that enough sample is taken to ensure that all components can be detected. In fact, when the goal of an analysis is to identify components present at trace levels, it may be desirable to discriminate against major components when sampling. In a quantitative analysis, however, the sample's composition must accurately represent the target population. The focus of this section, therefore, is on designing a sampling plan for a quantitative analysis.

Five questions should be considered when designing a sampling plan:

1. From where within the target population should samples be collected?
2. What type of samples should be collected?
3. What is the minimum amount of sample needed for each analysis?
4. How many samples should be analyzed?
5. How can the overall variance be minimized?

Each of these questions is considered below in more detail.

7B.1 Where to Sample the Target Population

Sampling errors occur when a sample's composition is not identical to that of the population from which it is drawn. When the material being sampled is homogeneous, individual samples can be taken without regard to possible sampling errors. Unfortunately, in most situations the target population is heterogeneous in either time or space. As a result of settling, for example, medications available as oral suspensions may have a higher concentration of their active ingredients at the bottom of the container. Before removing a dose (sample), the suspension is shaken to minimize the effect of this spatial heterogeneity. Clinical samples, such as blood or urine, frequently show a temporal heterogeneity. A patient's blood glucose level, for instance, will change in response to eating, medication, or exercise. Other systems show both spatial and temporal heterogeneities. The concentration of dissolved O_2 in a lake shows a temporal heterogeneity due to the change in seasons, whereas point sources of pollution may produce a spatial heterogeneity.

When the target population's heterogeneity is of concern, samples must be acquired in a manner that ensures that determinate sampling errors are insignificant. If the target population can be thoroughly homogenized, then samples can be taken without introducing sampling errors. In most cases, however, homogenizing the

target population is impracticable. Even more important, homogenization destroys information about the analyte's spatial or temporal distribution within the target population.

Random Sampling The ideal sampling plan provides an unbiased estimate of the target population's properties. This requirement is satisfied if the sample is collected at random from the target population.[3] Despite its apparent simplicity, a true **random sample** is difficult to obtain. Haphazard sampling, in which samples are collected without a sampling plan, is not random and may reflect an analyst's unintentional biases. The best method for ensuring the collection of a random sample is to divide the target population into equal units, assign a unique number to each unit, and use a random number table (Appendix 1E) to select the units from which to sample. Example 7.3 shows how this is accomplished.

random sample
A sample collected at random from the target population.

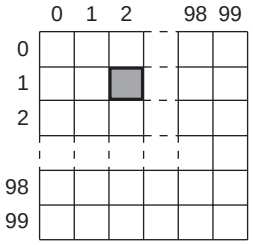
Example

EXAMPLE 7.3

To analyze the properties of a 100 cm × 100 cm polymer sheet, ten 1 cm × 1 cm samples are to be selected at random and removed for analysis. Explain how a random number table can be used to ensure that samples are drawn at random.

SOLUTION

As shown in the following grid, we divide the polymer sheet into 10,000 1 cm × 1 cm squares, each of which can be identified by its row number and its column number.



For example, the highlighted square is in row 1 and column 2. To pick ten squares at random, we enter the random number table at an arbitrary point, and let that number represent the row for the first sample. We then move through the table in a predetermined fashion, selecting random numbers for the column of the first sample, the row of the second sample, and so on until all ten samples have been selected. Since our random number table (Appendix 1E) uses five-digit numbers we will use only the last two digits. Let's begin with the fifth entry and use every other entry after that. The fifth entry is 65423 making the first row number 23. The next entry we use is 41812, giving the first column number as 12. Continuing in this manner, the ten samples are as follows:

Sample	Row	Column	Sample	Row	Column
1	23	12	6	93	83
2	45	80	7	91	17
3	81	12	8	45	13
4	66	17	9	12	92
5	46	01	10	97	52

A randomly collected sample makes no assumptions about the target population, making it the least biased approach to sampling. On the other hand, random sampling requires more time and expense than other sampling methods since a greater number of samples are needed to characterize the target population.

judgmental sampling
Samples collected from the target population using available information about the analyte's distribution within the population.

Judgmental Sampling The opposite of random sampling is selective, or **judgmental sampling,** in which we use available information about the target population to help select samples. Because assumptions about the target population are included in the sampling plan, judgmental sampling is more biased than random sampling; however, fewer samples are required. Judgmental sampling is common when we wish to limit the number of independent variables influencing the results of an analysis. For example, a researcher studying the bioaccumulation of polychlorinated biphenyls (PCBs) in fish may choose to exclude fish that are too small or that appear diseased. Judgmental sampling is also encountered in many protocols in which the sample to be collected is specifically defined by the regulatory agency.

systematic sampling
Samples collected from the target population at regular intervals in time or space.

Systematic Sampling Random sampling and judgmental sampling represent extremes in bias and the number of samples needed to accurately characterize the target population. **Systematic sampling** falls in between these extremes. In systematic sampling the target population is sampled at regular intervals in space or time. For a system exhibiting a spatial heterogeneity, such as the distribution of dissolved O_2 in a lake, samples can be systematically collected by dividing the system into discrete units using a two- or three-dimensional grid pattern (Figure 7.2). Samples are collected from the center of each unit, or at the intersection of grid lines. When a heterogeneity is time-dependent, as is common in clinical studies, samples are drawn at regular intervals.

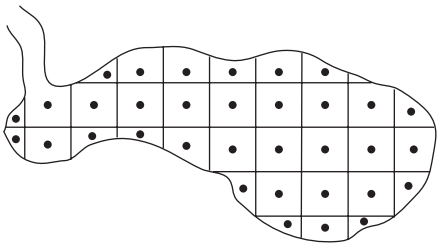


Figure 7.2
Example of a systematic sampling plan for collecting samples from a lake. Each solid dot represents a sample collected from within the sampling grid.

When a target population's spatial or temporal heterogeneity shows a periodic trend, a systematic sampling leads to a significant bias if samples are not collected frequently enough. This is a common problem when sampling electronic signals, in which case the problem is known as aliasing. Consider, for example, a signal consisting of a simple sine wave. Figure 7.3a shows how an insufficient sampling frequency underestimates the signal's true frequency.

Nyquist theorem
Statement that a periodic signal must be sampled at least twice each period to avoid a determinate error in measuring its frequency.

According to the **Nyquist theorem,** to determine a periodic signal's true frequency, we must sample the signal at a rate that is at least twice its frequency (Figure 7.3b); that is, the signal must be sampled at least twice during a single cycle or period. When samples are collected at an interval of Δt, the highest frequency that can be accurately monitored has a frequency of $(2\ \Delta t)^{-1}$. For example, if samples are collected every hour, the highest frequency that we can monitor is 0.5 h^{-1}, or a periodic cycle lasting 2 h. A signal with a cycling period of less than 2 h (a frequency of more than 0.5 h^{-1}) cannot be monitored. Ideally, the sampling frequency should be at least three to four times that of the highest frequency signal of interest. Thus, if an hourly periodic cycle is of interest, samples should be collected at least every 15–20 min.

systematic–judgmental sampling
A sampling plan that combines judgmental sampling with systematic sampling.

Systematic–Judgmental Sampling Combinations of the three primary approaches to sampling are also possible.[4] One such combination is **systematic–judgmental sampling,** which is encountered in environmental studies when a spatial or tempo-

ral distribution of pollutants is anticipated. For example, a plume of waste leaching from a landfill can reasonably be expected to move in the same direction as the flow of groundwater. The systematic–judgmental sampling plan shown in Figure 7.4 includes a rectangular grid for systematic sampling and linear transects extending the sampling along the plume's suspected major and minor axes.[5]

Stratified Sampling Another combination of the three primary approaches to sampling is judgmental–random, or **stratified sampling.** Many target populations are conveniently subdivided into distinct units, or strata. For example, in determining the concentration of particulate Pb in urban air, the target population can be subdivided by particle size. In this case samples can be collected in two ways. In a random sampling, differences in the strata are ignored, and individual samples are collected at random from the entire target population. In a stratified sampling the target population is divided into strata, and random samples are collected from within each stratum. Strata are analyzed separately, and their respective means are pooled to give an overall mean for the target population.

The advantage of stratified sampling is that the composition of each stratum is often more homogeneous than that of the entire target population. When true, the sampling variance for each stratum is less than that when the target population is treated as a single unit. As a result, the overall sampling variance for stratified sampling is always at least as good as, and often better than, that obtained by simple random sampling.

Convenience Sampling One additional method of sampling deserves brief mention. In **convenience sampling,** sample sites are selected using criteria other than minimizing sampling error and sampling variance. In a survey of groundwater quality, for example, samples can be collected by drilling wells at randomly selected sites, or by making use of existing wells. The latter method is usually the preferred choice. In this case, cost, expedience, and accessibility are the primary factors used in selecting sampling sites.

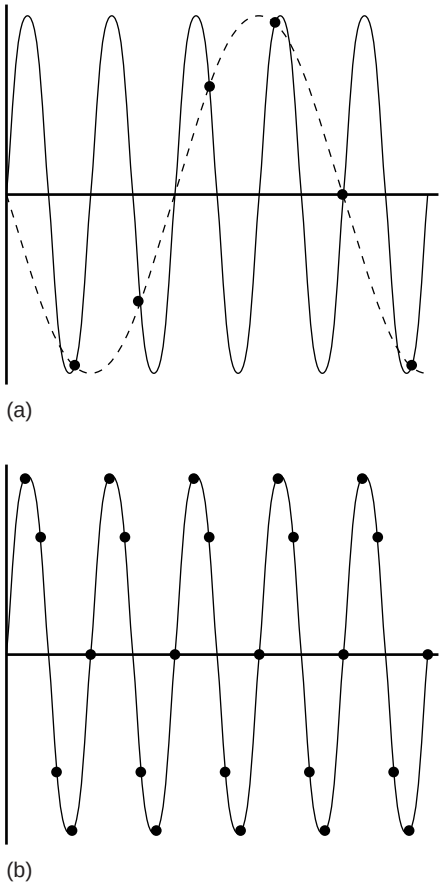


Figure 7.3
Effect of sampling frequency when monitoring periodic signals. In (a) the sampling frequency is 1.2 samples per period. The dashed line shows the apparent signal, while the solid line shows the true signal. In (b) a sampling frequency of five samples per period is sufficient to give an accurate estimation of the true signal.

stratified sampling
A sampling plan that divides the population into distinct strata from which random samples are collected.

convenience sampling
A sampling plan in which samples are collected because they are easily obtained.

grab sample
A single sample removed from the target population.

7B.2 What Type of Sample to Collect

After determining where to collect samples, the next step in designing a sampling plan is to decide what type of sample to collect. Three methods are commonly used to obtain samples: grab sampling, composite sampling, and in situ sampling. The most common type of sample is a **grab sample,** in which a portion of the target population is removed at a given time and location in space. A grab sample, therefore, provides a "snapshot" of the target population. Grab sampling is easily adapted to any of the sampling schemes discussed in the previous section. If the target population is fairly uniform in time and space, a set of grab samples collected at random can be used to establish its properties. A systematic sampling using grab samples can be used to characterize a target population whose composition varies over time or space.

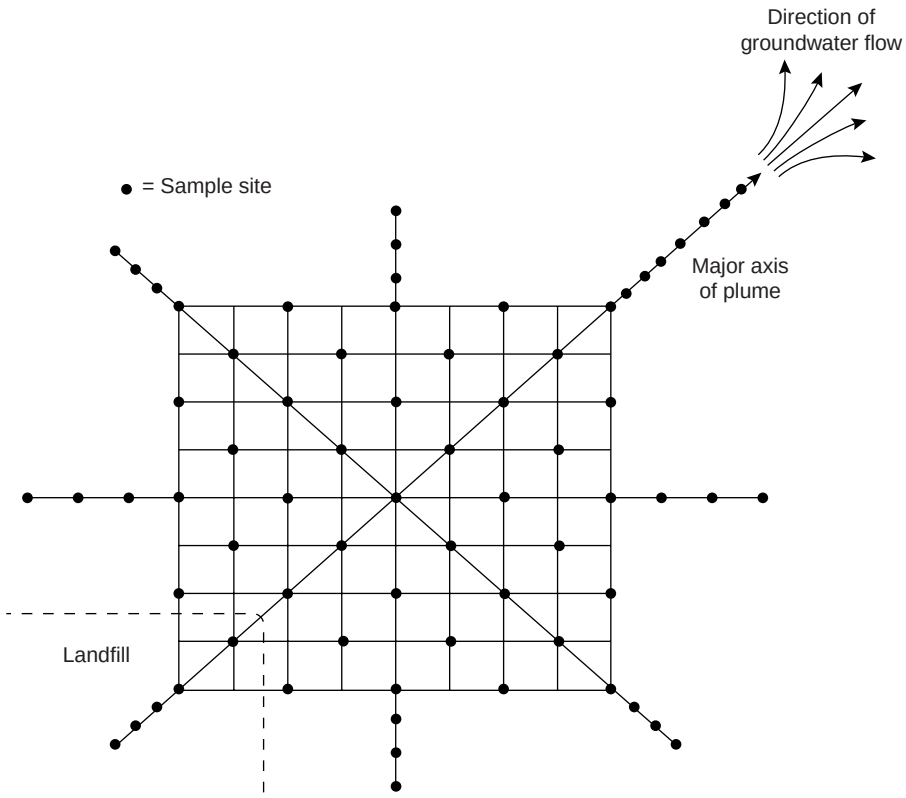


Figure 7.4
Systematic–judgmental sampling scheme for monitoring the leaching of pollutants from a landfill. Sites where samples are collected are represented by the solid dots.

composite sample
Several grab samples combined to form a single sample.

A **composite sample** consists of a set of grab samples that are combined to form a single sample. After thoroughly mixing, the composite sample is analyzed. Because information is lost when individual samples are combined, it is normally desirable to analyze each grab sample separately. In some situations, however, there are advantages to working with composite samples. One such situation is in determining a target population's average composition over time or space. For example, wastewater treatment plants are required to monitor and report the average composition of treated water released to the environment. One approach is to analyze a series of individual grab samples, collected using a systematic sampling plan, and average the results. Alternatively, the individual grab samples can be combined to form a single composite sample. Analyzing a single composite sample instead of many individual grab samples, provides an appreciable savings in time and cost. Composite sampling is also useful when a single sample cannot supply sufficient material for an analysis. For example, methods for determining PCBs in fish often require as much as 50 g of tissue, an amount that may be difficult to obtain from a single fish. Tissue samples from several fish can be combined and homogenized, and a 50-g portion of the composite sample taken for analysis.

in situ sampling
Sampling done within the population without physically removing the sample.

A significant disadvantage of grab samples and composite samples is the need to remove a portion of the target population for analysis. As a result, neither type of sample can be used to continuously monitor a time-dependent change in the target population. **In situ sampling,** in which an analytical sensor is placed directly in the target population, allows continuous monitoring without removing individual grab samples. For example, the pH of a solution moving through an industrial production line can be continually monitored by immersing a pH electrode within the solution's flow.

7B.3 How Much Sample to Collect

To minimize sampling errors, a randomly collected grab sample must be of an appropriate size. If the sample is too small its composition may differ substantially from that of the target population, resulting in a significant sampling error. Samples that are too large, however, may require more time and money to collect and analyze, without providing a significant improvement in sampling error.

As a starting point, let's assume that our target population consists of two types of particles. Particles of type A contain analyte at a fixed concentration, and type B particles contain no analyte. If the two types of particles are randomly distributed, then a sample drawn from the population will follow the binomial distribution.* If we collect a sample containing n particles, the expected number of particles containing analyte, n_A, is

$$n_A = np$$

where p is the probability of selecting a particle of type A. The sampling standard deviation is

$$s_s = \sqrt{np(1-p)} \qquad \textbf{7.3}$$

The relative standard deviation for sampling, $s_{s,r}$, is obtained by dividing equation 7.3 by n_A.

$$s_{s,r} = \frac{\sqrt{np(1-p)}}{np}$$

Solving for n allows us to calculate the number of particles that must be sampled to obtain a desired sampling variance.

$$n = \frac{1-p}{p} \times \frac{1}{s_{s,r}^2} \qquad \textbf{7.4}$$

Note that the relative sampling variance is inversely proportional to the number of particles sampled. Increasing the number of particles in a sample, therefore, improves the sampling variance.

Example

EXAMPLE 7.4

Suppose you are to analyze a solid where the particles containing analyte represent only 1×10^{-7}% of the population. How many particles must be collected to give a relative sampling variance of 1%?

SOLUTION

Since the particles of interest account for 1×10^{-7}% of all particles in the population, the probability of selecting one of these particles is only 1×10^{-9}. Substituting into equation 7.4 gives

$$n = \frac{1-(1\times 10^{-9})}{1\times 10^{-9}} \times \frac{1}{(0.01)^2} = 1 \times 10^{13}$$

Thus, to obtain the desired sampling variance we need to collect 1×10^{13} particles.

*See Chapter 4 to review the properties of a binomial distribution.

A sample containing 10^{13} particles can be fairly large. Suppose this is equivalent to a mass of 80 g. Working with a sample this large is not practical; but does this mean we must work with a smaller sample and accept a larger relative sampling variance? Fortunately the answer is no. An important feature of equation 7.4 is that the relative sampling variance is a function of the number of particles but not their combined mass. We can reduce the needed mass by crushing and grinding the particles to make them smaller. Our sample must still contain 10^{13} particles, but since each particle is smaller their combined mass also is smaller. If we assume that a particle is spherical, then its mass is proportional to the cube of its radius.

$$\text{Mass} \propto r^3$$

Decreasing a particle's radius by a factor of 2, for example, decreases its mass by a factor of 2^3, or 8. Instead of an 80-g sample, a 10-g sample will now contain 10^{13} particles.

Example

EXAMPLE 7.5

Assume that the sample of 10^{13} particles from Example 7.4 weighs 80 g. By how much must you reduce the radius of the particles if you wish to work with a sample of 0.6 g?

SOLUTION

To reduce the sample from 80 g to 0.6 g you must change its mass by a factor of

$$\frac{80}{0.6} = 133 \text{ times}$$

This can be accomplished by decreasing the radius of the particles by a factor of

$$x^3 = 133$$

$$x = 5.1$$

Decreasing the radius by a factor of approximately 5 allows you to decrease the sample's mass from 80 g to 0.6 g.

Treating a population as though it contains only two types of particles is a useful exercise because it shows us that the relative sampling variance can be improved by collecting more particles of sample. Furthermore, we learned that the mass of sample needed can be reduced by decreasing particle size without affecting the relative sampling variance. Both are important conclusions.

Few populations, however, meet the conditions for a true binomial distribution. Real populations normally contain more than two types of particles, with the analyte present at several levels of concentration. Nevertheless, many well-mixed populations, in which the population's composition is homogeneous on the scale at which we sample, approximate binomial sampling statistics. Under these conditions the following relationship between the mass of a randomly collected grab sample, m, and the percent relative standard deviation for sampling, R, is often valid.[6]

$$mR^2 = K_s \qquad \textbf{7.5}$$

where K_s is a sampling constant equal to the mass of sample producing a percent relative standard deviation for sampling of ±1%.* The sampling constant is evalu-

*Problem 8 in the end-of-chapter problem set asks you to consider the relationship between equations 7.4 and 7.5.