

Contents

Preface xi

CHAPTER 1 Introduction to Statistics and Quality Improvement 1

- 1.1 What Is Statistics? 1
- 1.2 Why Study Statistics? 2
- 1.3 Statistical Thinking: Understanding and Managing Variability 4
- 1.4 Variables, Types of Data, and Levels of Measurement 4
- 1.5 Operational Definitions 8
- 1.6 Sampling 10
- 1.7 Statistical and Spreadsheet Software 15
- 1.8 Introduction to Quality 16
- 1.9 A History of Quality and Productivity 18
- 1.10 Themes of Quality Management 21
- 1.11 The Connection between Quality and Statistics 26
 - Whitney Gourmet Cat Food Company Case 28
- Appendix 1.1 Basics of the Windows User Interface 31
- Appendix 1.2 Introduction to Microsoft Excel 33
- Appendix 1.3 Introduction to MINITAB 41

CHAPTER 2 Tables and Charts 44

- 2.1 Introduction and the History of Graphics 44
- 2.2 Some Tools for Studying a Process: Process Flow Diagrams and Cause-and-Effect Diagrams 45
- 2.3 The Importance of the Time-Order Plot 52
- 2.4 Tables and Charts for Numerical Data 56
- 2.5 Checksheets and Summary Tables 69
- 2.6 Concentration Diagrams 70
- 2.7 Graphing Categorical Data 72
- 2.8 Tables and Charts for Bivariate Categorical Data 78
- 2.9 Graphical Excellence 81
 - Whitney Gourmet Cat Food Company Case 93
- Appendix 2.1 Using Microsoft Excel for Tables and Charts 96
- Appendix 2.2 Using MINITAB for Tables and Charts 101

CHAPTER 3 Describing and Summarizing Data 103

- 3.1 Introduction: What's Ahead 103
- 3.2 Measures of Central Tendency, Variation, and Shape 104
- 3.3 The Box-and-Whisker Plot 123

Whitney Gourmet Cat Food Company Case 133

Appendix 3.1 Using Microsoft Excel for Descriptive Statistics 133

Appendix 3.2 Using MINITAB for Descriptive Statistics 134

CHAPTER 4 Probability and Discrete Probability Distributions 135

- 4.1 Introduction 135
- 4.2 Some Rules of Probability 138
- 4.3 The Probability Distribution 144
- 4.4 The Binomial Distribution 150
- 4.5 The Hypergeometric Distribution 158
- 4.6 The Negative Binomial and Geometric Distributions 162
- 4.7 The Poisson Distribution 167
- 4.8 Summary and Overview 172

Appendix 4.1 Using Microsoft Excel for Probability and Probability Distributions 176

Appendix 4.2 Using MINITAB for Probability and Probability Distributions 179

CHAPTER 5 Continuous Probability Distributions and Sampling Distributions 180

- 5.1 Introduction to Continuous Probability Distributions 180
- 5.2 The Uniform Distribution 181
- 5.3 The Normal Distribution 184
- 5.4 The Standard Normal Distribution As An Approximation to the Binomial and Poisson Distributions 196
- 5.5 The Normal Probability Plot 199
- 5.6 The Lognormal Distribution 206
- 5.7 The Exponential Distribution 209
- 5.8 The Weibull Distribution 211
- 5.9 Sampling Distribution of the Mean 213
- 5.10 Sampling Distribution of the Proportion 218
- 5.11 Summary 220

Whitney Gourmet Cat Food Company Case 224

Appendix 5.1 Using Microsoft Excel for Continuous Probability Distributions and Sampling Distributions 225

Appendix 5.2 Using MINITAB for Continuous Probability Distributions and Sampling Distributions 228

CHAPTER 6 Statistical Process Control Charts I: Basic Concepts and Attributes Charts 230

- 6.1 Introduction to Control Charts and Their Applications 230
- 6.2 Introduction to the Theory of Control Charts 233
- 6.3 Introduction to Attributes Control Charts 239
- 6.4 np and p charts 240
- 6.5 Area of Opportunity Charts (c Charts and u Charts) 256
- 6.6 Summary 267

The Whitney Gourmet Cat Food Company Case 279

The Mariphil Farms Dairy Case 282

Appendix 6.1 Using Microsoft Excel for Attribute Control Charts 284

Appendix 6.2 Using MINITAB for Attribute Control Charts 285

CHAPTER 7 Statistical Process Control Charts II: Variables Control Charts 287

- 7.1 Introduction to Variables Control Charts 287
- 7.2 Rational Subgroups and Sampling Decisions 288
- 7.3 Control Charts for Central Tendency (\bar{X} Charts) and Variation (R and s Charts) 290
- 7.4 Control Charts for Individual Values (X Charts) 311
- 7.5 Special Considerations with Variables Charts 325
- 7.6 The Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) Charts 326
- 7.7 Process Capability 332
- 7.8 Summary 339

The Whitney Gourmet Cat Food Company Case 351

Harnswell Sewing Machine Company Case 353

Antelope Pass Mining Company Case 357

Appendix 7.1 Using Microsoft Excel for Variables Control Charts 358

Appendix 7.2 Using MINITAB for Variables Control Charts 363

CHAPTER 8 Estimation Procedures 365

- 8.1 Introduction 365
- 8.2 Properties of Estimators 365
- 8.3 Confidence Interval Estimation of the Mean 367
- 8.4 Confidence Interval Estimation for the Variance 376
- 8.5 Prediction Interval Estimate for a Future Individual Value 379
- 8.6 Tolerance Intervals 381
- 8.7 Confidence Interval Estimation for the Proportion 384

- 8.8 Summary 386
 - Whitney Gourmet Cat Food Company Case 388
- Appendix 8.1 Using Microsoft Excel for Confidence Interval Estimation 389
- Appendix 8.2 Using MINITAB for Confidence Interval Estimation 390

CHAPTER 9 Introduction to Hypothesis Testing 392

- 9.1 Introduction 392
- 9.2 Basic Concepts of Hypothesis-Testing 393
- 9.3 One-Sample Tests for the Mean 398
- 9.4 t Test for the Difference between the Means of Two Independent Groups 413
- 9.5 Testing for the Difference between Two Variances 422
- 9.6 The Repeated Measures or Paired t Test 427
- 9.7 Chi-Square Test for the Differences among Proportions in Two or More Groups 433
- 9.8 χ^2 Test of Hypothesis for the Variance or Standard Deviation (Optional Topic) 447
- 9.9 Wilcoxon Rank Sum Test for the Difference between Two Medians (Optional Topic) 450
- 9.10 Summary 457
 - Whitney Gourmet Cat Food Company Case 461
- Appendix 9.1 Using Microsoft Excel for Hypothesis Testing 462
- Appendix 9.2 Using MINITAB for Hypothesis Testing 467

CHAPTER 10 The Design of Experiments: One Factor and Randomized Block Experiments 469

- 10.1 Introduction and Rationale 469
- 10.2 Historical Background 470
- 10.3 The Concept of Randomization 471
- 10.4 The One-Way Analysis of Variance (ANOVA) 472
- 10.5 The Randomized Block Model 486
- 10.6 Kruskal-Wallis Rank Test for Differences in c Medians (Optional Topic) 497
 - Whitney Gourmet Cat Food Company Case 508
- Appendix 10.1 Using Microsoft Excel for the Analysis of Variance 509
- Appendix 10.2 Using MINITAB for the Analysis of Variance 511

CHAPTER 11 The Design of Experiments: Factorial Designs 513

- 11.1 Two-Factor Factorial Designs 513
- 11.2 Factorial Designs Involving Three or More Factors 525

- 11.3 The Fractional Factorial Design 541
- 11.4 The Taguchi Approach 550
- 11.5 Summary and Overview 557
 - Whitney Gourmet Cat Food Company Case 563
- Appendix 11.1 Using Microsoft Excel for the Two-Factor Factorial Design 565
- Appendix 11.2 Using MINITAB for Factorial Designs 566

CHAPTER 12 Simple Linear Regression and Correlation 567

- 12.1 Introduction 567
- 12.2 Types of Regression Models 568
- 12.3 Determining the Simple Linear Regression Equation 570
- 12.4 Measures of Variation in Regression and Correlation 577
- 12.5 Assumptions of Regression and Correlation 581
- 12.6 Residual Analysis 582
- 12.7 Inferences About the Slope 587
- 12.8 Confidence and Prediction Interval Estimation 591
- 12.9 Pitfalls in Regression and Ethical Issues 596
- 12.10 Computations in Simple Linear Regression 599
- 12.11 Correlation—Measuring the Strength of the Association 603
 - Whitney Gourmet Cat Food Company Case 612
- Appendix 12.1 Using Microsoft Excel for Simple Linear Regression and Correlation 613
- Appendix 12.2 Using MINITAB for Simple Linear Regression and Correlation 615

CHAPTER 13 Multiple Regression 616

- 13.1 Developing the Multiple-Regression Model 616
- 13.2 Residual Analysis for the Multiple-Regression Model 624
- 13.3 Testing for the Significance of the Multiple-Regression Model 626
- 13.4 Inferences Concerning the Population Regression Coefficients 628
- 13.5 Testing Portions of the Multiple-Regression Model 631
- 13.6 The Quadratic Curvilinear Regression Model 638
- 13.7 Dummy-Variable Models 646
- 13.8 Using Transformations in Regression Models 651
- 13.9 Collinearity 654
- 13.10 Model-Building 655
- 13.11 Pitfalls in Multiple Regression 664
- Appendix 13.1 Using Microsoft Excel for Multiple-Regression Models 668
- Appendix 13.2 Using MINITAB for Multiple Models Regression 670

Appendices

A. Tables A-1

- A.1 Table of Random Numbers A-2
- A.2 The Standard Normal Distribution A-4
- A.3 Control Chart Factors A-5
- A.4 Critical Values of t A-6
- A.5a K_2 Factors for Normal Distributions A-8
- A.5b K_1 Factors for One-Sided Tolerance Intervals for Normal Distributions A-9
- A.6 Critical Values of χ^2 A-10
- A.7 Critical Values of F A-11
- A.8 Lower and Upper Critical Values T_1 of Wilcoxon Rank Sum Test A-15
- A.9 Critical Values of the Studentized Range Q

B. Statistical Forms B-1

- B.1 Normal Probability Paper B-1
- B.2 Sample form for Data Collection and Construction of Attribute Control Charts B-2

C. Documentation for the Data Files C-1

D. Installing the PHStat Microsoft Excel Add-In D-1

E. Answers to Selected Odd Problems E-1

Index I-1

Preface

Introduction

The primary questions that must be answered when a new statistics text for engineers and scientists is written relate to the issue of the contribution of the textbook to the pedagogy of teaching statistics to this audience of students and to how the text will differ from the many texts that are already available. These questions can be answered for the proposed text only in the context of recommendations that have been made as the result of a 1984 conference on the statistical education of engineers [Hogg(1985)] and a 1993 Quality Engineering Workshop [Hogg(1994)]. Among the recommendations made was that engineers need to appreciate the following statistical concepts:

1. omnipresence of variability;
2. the use of simple graphical tools such as run charts, histograms, scatter plots, probability plots, residual plots, and control charts;
3. basic concepts of statistical inference;
4. the importance and essentials of carefully planned design of experiments;
5. the philosophies of Shewhart, Deming, and other practitioners concerning the quality of products and services.

The Hogg(1994) article proposed a core course of topics for engineering students. This proposed text is based on the curriculum model presented in that article.

Educational Philosophy

In our many years of teaching introductory statistics courses to students majoring in a wide variety of disciplines, we have continually searched for ways to improve the teaching of these courses. Over the years, our vision has come to include the following:

1. Students need a frame of reference when learning a subject, especially one that is not their major. This frame of reference for engineering and science students should be applications to the various areas of engineering and the sciences. The discussion of each statistical topic should include references to at least one of these areas of application.
2. Virtually all the students taking introductory statistics courses for engineers and scientists are majoring in areas other than statistics. Introductory courses should, therefore, focus on the underlying principles that are important for nonstatistics majors.
3. The use of spreadsheet and/or statistical software should be integrated into all aspects of the introductory statistics course. The reality that exists in the workplace is that spreadsheet software (and sometimes statistical software)

is typically available on one's computer desktop. Our teaching approach needs to recognize this reality and make our courses more consistent with the workplace environment.

4. Textbooks that use software must provide detailed instructions that maximizes the student's ability to use the software with a minimum risk of failure.
5. Instruction for each topic should focus on (1) the application of the topic to an area of engineering or the sciences, (2) the interpretation of results, (3) the presentation of assumptions, (4) the evaluation of the assumptions, and (5) the discussion of what should be done if the assumptions are violated. These points are particularly important in regression, forecasting and in hypothesis testing. Although the illustration of some computations is inevitable, the focus on computations should be minimized.
6. Both classroom examples and homework exercises should relate to actual or realistic data as much as possible. Introductory courses should avoid an over-concentration on one topic area and instead provide breadth of coverage of a variety of statistical topics. This will help students avoid the "I can't see the forest for the trees" syndrome.

The main features of this proposed text are summarized in the following sections.

Main Feature: Emphasis on Data Analysis and Interpretation of Computer Output

The personal computer revolution has dramatically changed how information is analyzed in the workplace and how statistics should be taught in the classroom. In this text, we take the position that the use of computer software in the form of a spreadsheet application such as Microsoft Excel or a statistical package such as MINITAB is an integral part of learning statistics. We emphasize analyzing data, interpreting the output from Microsoft Excel and MINITAB, and explaining how to use this software while reducing the emphasis on computation. In order to carry out our approach, we have integrated this output into the fabric of the text. For example, our coverage of tables and charts in Chapter 2 focuses on the interpretation of various charts, not on their construction by hand. In Chapter 9 on hypothesis testing, we have made sure to include extensive computer output so that the p -value approach can be used. The presentation of simple linear regression in Chapter 12, assumes that software such as Microsoft Excel or MINITAB will be used, and thus our focus is on the interpretation of the output, not on hand calculations (which have been placed in a separate section of the chapter).

Main Feature: Problems, Case Studies, and Team Projects

"Learning" results from "doing." This text provides the student with the opportunity to select from many problems (most with multiple parts) presented at the ends of sections as well as at the ends of chapters. Most of these problems use real data and apply to realistic situations in various fields of engineering and the sciences. Students can aid their comprehension by engaging in multiple hands-on exercises as detailed below.

- The end-of-section problems give the students the opportunity to reinforce what they have just learned.
- The chapter review problems included at the end of each chapter are based on the concepts and methods learned throughout the chapter.

- Answers to Selected Odd-Numbered Problems appear at the end of the text.
- Detailed Case Studies are included at the end of several chapters.
- The *Whitney Gourmet Cat Food Company* case study is included at the end of most chapters, as an integrating theme.

Main Feature: Appendices on Using Microsoft Excel and MINITAB

Rather than rely on the supplementary manuals, that accompany statistical software packages, it is a much better pedagogical approach to provide an explanation of how the software is used in the text while employing the in-chapter examples. Detailed appendices are included at the end of all chapters that explain how to use MINITAB, the most popular statistical software for introductory business statistics, and Microsoft Excel, the dominant spreadsheet package. In addition, an appendix is provided after Chapter 1 that explains the basics of the Windows operating environment.

Main Feature: Statistics Add-In for Microsoft Excel—PHStat

The CD-ROM that accompanies the text includes the PHStat Statistics add-in for Microsoft Excel that facilitates its use in introductory statistics courses. Although Microsoft Excel is a spreadsheet package, it contains features that enable it to perform statistical analysis for many of the topics in this text. In some cases, however, such analyses are cumbersome in the off-the-shelf version of Microsoft Excel. The PHStat statistics add-in provides a custom menu of choices that leads to dialog boxes which enable users to make entries and selections to perform specific analyses. PHStat minimizes the work associated with setting up statistical solutions in Microsoft Excel by automating the creation of spreadsheets and charts. PHStat, along with Microsoft Excel's Data Analysis tool, now allows users to perform statistical analyses on virtually all topics covered in this text.

Main Feature: Pedagogical Aids

Numerous features designed to create a more stimulating learning environment throughout the text include:

- conversational writing style;
- a *Using Statistics* example that illustrates the application of at least one of the statistical methods covered in each chapter in engineering and the sciences;
- real data for many of the examples and problems;
- exhibit boxes that highlight important concepts;
- comment boxes that focus on assumptions of statistical methods;
- problem sets with varied levels of difficulty and complexity;
- key terms;
- chapter opening quotes from a philosopher, historical figure, well-known statistician or from literature;
- chapter ending problems that begin with *Checking Your Understanding* problems that require students to demonstrate their understanding of concepts;
- explanation and illustration of statistical tables;
- side notes in which additional material appears adjacent to where it is referenced.

Main Feature: Statistical Topics Covered

The text focuses on such topics as tables and charts (Chapter 2), descriptive statistics (Chapter 3), control charts (Chapters 6 and 7), experimental design (Chapters 10 and 11), regression (Chapters 12 and 13), and statistical inference (Chapters 8 and 9). This emphasis is consistent with the recommendations presented by Hogg (1994).

Perhaps the important statistical method used by engineers in industry is experimental design. Simply stated, engineers need to know how to conduct experiments where multiple factors are varied. Thus, in addition to coverage of one- and two-factor designs, this text discusses the concept of interaction in depth. Further, it provides coverage of factorial and fractional-factorial designs, using both a graphical approach and a confirmatory hypothesis-testing approach. In addition, the contributions of the Japanese engineer Genichi Taguchi are introduced.

By providing this comprehensive coverage of quality and experimental design, the text provides an orientation that allows the presentation of statistical tools in an organizational context, instead of in isolation. The goal is for students to learn not just how to use the tools but why and how statistical methods are useful in a wide variety of industrial settings. A portion of Chapter 1 is devoted to quality management, including both key themes and the contribution of individuals such as W. Edwards Deming, Joseph Juran, and Walter Shewhart.

Main Feature: Full Supplement Package

The supplement package that accompanies this text includes:

- **Instructor's Solution Manual**—written by M&N Toscano. This solutions manual is enriched with extra detail in the problem solutions and many Excel solutions. ISBN 0-13-027423-2.
- **Student Solutions Manual**—written by M&N Toscano. To order this supplement for your students, use ISBN 0-13-028681-8.

About the World Wide Web

The text has a home page on the World Wide Web with an address of <http://www.prenhall.com/levine>. There is a separate home page on the World Wide Web for the PHStat add-in, <http://www.prenhall.com/phstat> that provides user assistance and periodic updates.

Acknowledgements

We are extremely grateful to the many organizations and companies that allowed us to use their data in developing problems and examples throughout the text. We would like to thank American Cyanamid Company, American Society for Testing and Materials, Biometrika, Environmental Progress, Graphics Press, Journal of Energy Resources Technology, Journal of Engineering for Industry, Journal of Structural Engineering, Journal of the Minerals, Metals and Materials Society, Journal of Water Resources Planning and Management, New England Journal of Medicine, Newsday, Noise Control Engineering Journal, Philosophical Transactions of the Royal Society,

Quality and Reliability Engineering International, Quality Engineering, Quality Progress, Technometrics, The American Statistician, and The Free Press.

We would also like to express our gratitude to David Cresap, University of Portland, Dr. C. H. Aikens, The University of Tennessee—Knoxville, and Dr. Robert L. Armacost, University of Central Florida for their constructive comments during the writing of this text.

We offer special thanks to Kathy Boothby Sestak, Joanne Wendelken, Ann Heath, and Gina Huck of the editing team at Prentice Hall, and to Bob Walters our Production Editor. Thanks also to Brian Baker for his copyediting and M&N Toscano for their accuracy checking.

References

1. Hogg, R. V.; "A Core in Statistics for Engineering Students," *The American Statistician*, 48 (1994), 285–287
2. Hogg, R. V. et al.; "Statistical Education for Engineers: An Initial Task Force Report," *The American Statistician*, 39 (1985), 168–175

Concluding Remarks

We have gone to great lengths to make this text both pedagogically sound and error-free. If you have any suggestions, note material requiring clarification, or find potential errors, please contact us at **David_Levine@BARUCH.CUNY.EDU** or **DavidMLevine@MSN.COM**

David M. Levine

Patricia P. Ramsey

Robert K. Smidt

Introduction to Statistics and Quality Improvement

"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

H. G. WELLS



1.1 WHAT IS STATISTICS?

The term *statistics* can have many meanings. It is sometimes used as a synonym for numerical information. Such numerical values are often referred to as data or statistics. Examples of such statistics are grades on student exams, measures on a Richter scale, the diameter of a ball bearing, the amount of reactant in a chemical experiment, and the temperature increases in a supercomputer. When used in this context, the term statistics is a plural noun. The term statistics can also be used as a singular noun and, as such, represents a body of knowledge that enables one to do the following:

1. draw useful conclusions from numerical information;
2. make decisions in a rational and objective way;
3. predict and control events;
4. strive for excellence and increase quality and productivity.

The field of statistics is sometimes categorized into two areas. The first area is referred to as *descriptive statistics*.

Descriptive statistics consists of methods dealing with the collection, tabulation, summarization, and presentation of data.

These methods describe the various aspects of a data set. Descriptive statistical methods have their beginnings in the inventories kept by early civilizations, such as the Babylonians, Egyptians, and Chinese. For example, the Old Testament of the Bible refers to the numbering or counting of the people of Israel and to the casting of lots for selection by chance, and the Romans kept careful counts of people, possessions, and wealth in the territories they conquered. Similarly, the Domesday Book of the late eleventh century enumerated the lands and wealth of England. The Middle Ages also saw the growth of governments and religious institutions and their recording of births, deaths, and marriages. These early methods were primarily lists and counts kept for purposes of taxation and military conscription.

Although statistical methods developed throughout history in many different cultures, modern statistical concepts are considered to have developed in Europe in the late seventeenth century with the growth of mathematics and probability theory. (See Reference 10.) Theories of probability or chance have their historical roots in games of chance. By the seventeenth century, interest in gambling and the development of mathematical methods combined and resulted in early rules of probability. The development of theories of probability led to the inception of the second major field of statistics, called *inferential statistics*, in the beginning of the twentieth century. Statisticians such as Pearson, Fisher, Neyman, Wald, and Tukey pioneered in the development of the methods of inferential statistics, which are widely applied in so many fields today.

Inferential statistics consists of methods that permit one to reach conclusions and make estimates about populations based upon information from a sample.

Populations and samples are defined as follows:

A **population** or **universe** consists of all members of a class or category of interest. Its size is usually denoted by “*N*.”

A **sample** is some portion or subset of a population. Its size is usually denoted by “*n*.”

If every member of a population is evaluated, a **census** has been performed, and any summary value of all of the individual measurements is called a **parameter**. If only a subset or a sample of a population has been evaluated, any summary value of such measurements is called a **statistic**. Inferential statistics, therefore, involves using sample statistics to estimate population parameters.

A **census** is an enumeration or evaluation of every member of a population.
A **parameter** is a summary measure of the individual observations made in a census of an entire population.
A **statistic** is a summary measure of the individual observations made by evaluation of a sample.

1.2 WHY STUDY STATISTICS?

More than a half-century ago, Walter Shewhart, the father of statistical quality control, said that

“The long range contribution of statistics depends not so much upon getting a lot of highly trained statisticians into industry as it does in creating a statistically minded generation of physicists, chemists, engineers and others who will in any way have a hand in developing and directing the production processes of tomorrow.”

Engineers and scientists, as individuals who are continually applying the scientific method to their field of endeavor, are also constantly dealing with numerical information that needs to be analyzed. It is from this perspective that we should

consider why engineers and scientists need to know about statistics. Among the reasons for learning statistics are:

- 1. to know how to present and describe numerical information properly;
- 2. to know how to draw conclusions about large populations from sample information only;
- 3. to know how to improve processes and how to engineer quality into a product;
- 4. to design experiments to learn more about the natural world and to model physical relationships;
- 5. to know how to obtain reliable forecasts or predictions of real-world responses.

A road map of this text from the perspective of these five reasons for learning statistics is presented in Figure 1.1. From this road map, we observe that the first three chapters include coverage of methods involved in the collection, presentation, and description of information. Chapters 4 and 5 provide thorough coverage of the concepts of probability, probability distributions, and sampling distributions, so that subsequently the reader will learn how to draw conclusions about large populations based only upon information obtained from samples (which are taught in Chapters 8 and 9). Chapters 6 and 7 contain coverage of statistical applications in quality and productivity that are essential for process improvement. Chapters 10 and 11 discuss the fundamental principles of experimental design, which are critical to the proper use of the scientific method. The text concludes with Chapters 12 and 13, which focus

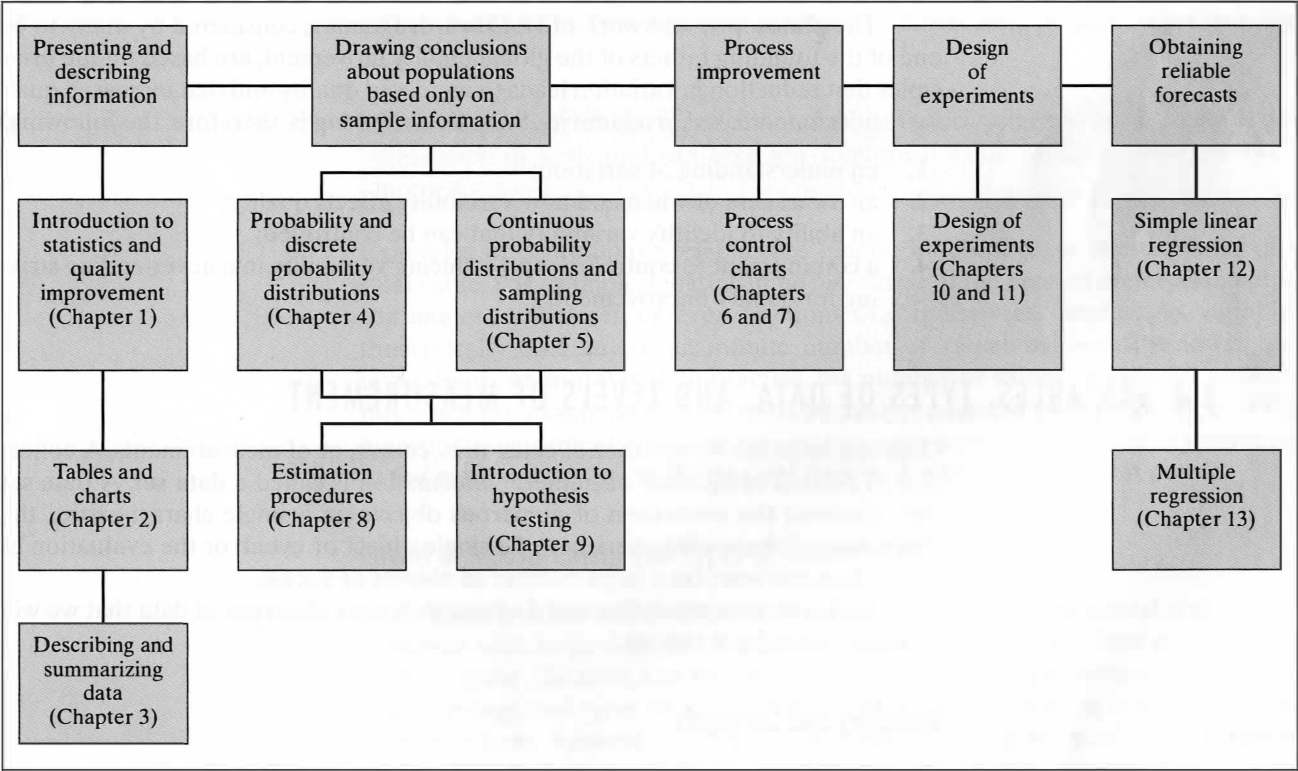


Figure 1.1 A Road Map of the Text

on regression and multiple-regression analysis that provides methods for obtaining forecasts.



1.3 STATISTICAL THINKING: UNDERSTANDING AND MANAGING VARIABILITY

Statistical thinking starts with an awareness and understanding that no two things are exactly alike and that variability is inherent in all things. Statistical thinking is the ability to identify, quantify, reduce, and control the kinds of variability that reduce the quality of our performance, of the objects we make, and of our actions as individuals, groups, and societies. Variability is not inherently undesirable. It is essential to variety and creativity and often enriches our lives. Ours would be a much poorer world if all people were identical and every artist painted exactly the same picture; however, in areas where uniformity, consistency, and dependability are essential or desirable, we must recognize that variability reduces quality and hampers excellence.

To function in the technological world, consistency is essential. Without consistency and predictability, interacting with people and objects would result in chaos. Research has shown that consumers value consistency in products and services and associate dependability and reliability with quality. Recent experience has shown that brands of automobiles perceived to be dependable and reliable have captured an ever-increasing share of the world market. Industry standards calling for uniformity of design and manufacture of various parts and products, such as light bulbs, automobile tires, screw threads, and computer disks, not only make high-volume production and distribution possible but also permit interchangeability of parts and simplify daily life and business.

The philosophy and work of W. Edwards Deming, considered by many to be one of the founding fathers of the global quality movement, are based on the principles that reduction in variation leads to increased quality and that increased quality results in increased productivity. **Statistical thinking** is, therefore, the following:

1. an understanding of variation;
2. an awareness of when and how variability affects quality;
3. an ability to identify variability that can be controlled;
4. a commitment to controlling and reducing variability in a never-ending striving for quality improvement.



1.4 VARIABLES, TYPES OF DATA, AND LEVELS OF MEASUREMENT

Data can be in the form either of categories, counts, or of measurements. A collection of related categorical or numeric information is called a **data set**. A data set can represent the evaluation of numerous objects on a single characteristic, the evaluation of many characteristics of a single object or event, or the evaluation of many characteristics on a large number of objects or events.

In this section, we define and distinguish among the types of data that we will be using throughout this text.

Variables and Constants

A **variable** can be defined as a characteristic that can take different values.

Examples of variables are measures of temperature, height, weight, cost, and gender.

Constants are characteristics that have values that do not change.

Examples of constants are pi (π), the ratio of the circumference of a circle to its diameter ($\pi = 3.14159 \dots$), and e , the base of the natural (or *Napierian*) logarithms ($e = 2.71828 \dots$).

Types of Variables

Different types of variables produce different types of data. **Categorical variables** yield values that can be placed only into categories or classes. Examples are gender, species, causes of a shutdown of a nuclear plant, type of cancer, and brand of computer owned. **Numerical variables** yield numerical values. One way of categorizing numerical variables is to consider them to be either *discrete* or *continuous*.

Discrete variables are *countable*, in that there is a gap between each possible value. Most often, these variables result from a process of counting and are thus restricted to whole numbers or integer values.

Discrete variables measure countable whole units, such as the number of defective welds in a length of pipe, the number of logic errors in a computer program, or the number of vehicles crossing a bridge in a 24-hour period. Because they are numerical counts of things, discrete variables usually answer the question, how many? Such counts are reported in whole numbers and, as such, represent discrete data.

Continuous variables represent numerical measurements on a *continuous* dimension or scale and can take any numerical value within a continuum or interval.

Examples of continuous variables are length, weight, age, or the diameter of a steel cable. For example, length can be measured to the nearest meter, decimeter, centimeter, millimeter, or even fractions of a millimeter. Continuous variables theoretically take any of an infinite number of values between the lower and upper limits of an interval. In practice, the number of values obtained is not infinite since the exactness of such measurements is limited by the precision of the measuring instruments used or conventions such as rounding monetary amounts to the nearest cent, age to the nearest year, or weight to the nearest pound.

Levels of Measurement and Types of Measurement Scales

In addition to considering variables as either categorical or numerical, data can be classified with respect to the degree to which it possesses the four properties of identity, order, distance, and an absolute zero point. These properties result in four widely recognized types of scales of measurement. They can be ordered in terms of power from weakest to strongest, so they are often referred to as **levels of measurement**. These levels of measurement are represented by the *nominal*, *ordinal*, *interval*, and *ratio* scales.

Nominal Scales

Nominal scales identify, name, classify, or categorize objects or events with no natural ordering of categories.

In a nominal scale objects or events are evaluated on a characteristic and classified into a category. Variables such as gender, type of computer, blood type, and species are examples of categorical variables. Each of these identifies a category into which individuals fall, so they are nominal variables. Additional examples of nominal variables are type of transmission in an automobile (automatic or manual), type of fuel used for heating a home, or type of electric wire used. The categories used when evaluating an individual or object on a nominal scale should be *mutually exclusive* and *collectively exhaustive*.

Mutually exclusive means that each item classified falls into one and only one category.

Collectively exhaustive means that there is a complete set of categories available, so that all items can be classified.

Nominally scaled variables are the weakest kind of data because they identify only the category into which the item being classified belongs. Because there is no logical or inherent order to the categories, the only analyses possible with this type of scale involves counts or proportions of items that fall into the various categories.

Ordinal Scales

Ordinal scales, like nominal scales, identify, name, classify, or categorize objects or events, but they have the additional property of a logical or natural order to the categories or values.

Examples of ordinal scales are tire rating, college class (freshman, sophomore, junior, and senior), or automobile size (subcompact, compact, intermediate, full-size). Any time that individuals or objects are *ranked* on some dimension, an ordinal scale of measurement is being used. Although ordinal scales are more powerful than nominal scales, because they have the properties of both identity and order, they are still a fairly weak form of measurement, because the distances between ordered categories do not represent equal intervals. Ordinal scales permit comparison of scores or categories in terms of smaller and larger, higher and lower, or from best to worst. They do not permit comparisons in terms of how much larger, lower, or better.

Interval Scales

Interval scales identify, have ordered values, and have the additional property of equal distances or intervals between scale values.

Interval scales identify, name, classify, or categorize as do nominal and ordinal scales and have ordered categories or values like ordinal scales. They have an important additional property: *equal distances* or *intervals* between scores or scale values. For example, temperature measured on either the Celsius or Fahrenheit

scales represents interval-scaled data. A one-degree difference on the Celsius scale represents the same quantity any place on the scale. A ten-degree increase in temperature from 20°C to 30°C represents the same temperature change as a ten-degree increase from 80°C to 90°C. Because intervals between scores are equal, scores can be compared based on the size of the differences between them. They can be added to and subtracted from one another and, in this manner, new variables can be created.

It is interesting to note that scores on classroom exams and standardized achievement tests such as the SAT, GRE, and GMAT are, at best, interval-scale measurements. A score of zero on an exam does not mean that a student knows nothing or has a complete absence of knowledge. It is, therefore, not reasonable to claim that one student is twice as knowledgeable as another is because he or she happened to earn an exam grade twice as high as the other does.

Ratio Scales

Ratio scales identify, order, represent equal distances between score values, and have an absolute zero point.

Ratio scales have all of the properties of nominal, ordinal, and interval scales, and they have an additional property: an absolute zero point. Sometimes this is referred to as a real zero point. This means that a scale value or score of zero represents a complete absence of the characteristic being measured. If temperature is measured on the Kelvin scale, a value of zero indicates a complete absence of temperature and a situation in which it would be impossible to become colder. Anyone who has spent time outdoors in winter in northern latitudes can attest to the fact that if the temperature is 0° as measured by either the Fahrenheit or Celsius scales, it can still become a good deal colder. This is why these scales are considered interval scales; they do not possess a real zero point.

Other examples of ratio scales are time, voltage, speed, electrical resistance, and diameter. Ratio scales have equal intervals between scale values like interval scales, so all the arithmetic operations that can be performed on interval data can be used with ratio data. In addition, the property of absolute zero permits one to make ratios of values. For example, a rod that is 3 feet long can be compared to a rod that is 6 feet long by making a ratio of their measurements. We could say that the 3-foot rod is half ($3/6 = 1/2$) the length of the 6-foot rod, or that the 6-foot rod is twice ($6/3 = 2$) the length of the 3-foot rod. This ability to make ratios of scores to compare measurements permits even more sophisticated forms of analysis. Ratio data are, therefore, considered to be the highest form of data available.

It is possible to transform data from a higher-level scale of measurement into a measure on a lower-level scale. For example, we could take a ratio scale such as length and collapse it into an ordinal scale by transforming the numerical measurements of length into ranks from longest to shortest, or we could collapse it further by classifying the individual lengths as short, medium, or long. Transformations in the opposite direction from a weak level of measurement to a higher level of measurement are not possible.

Attribute Measures and Variables Measures

Engineers and managers working in the area of quality improvement (discussed in Chapters 6 and 7) often use the terminology *attribute measures* and

variables measures to represent two classes of measurement. **Attribute measures** either are nominal variables based on counts or involve proportions calculated from discrete data. Examples of attribute measures are the number of defective or nonconforming items, the proportion of defective items, or the number or proportion of times an event occurs. **Variables measures** are continuous variables such as time measured in minutes, height or length measured in inches or centimeters, weight measured in pounds or kilograms, temperature measured in degrees Celsius or Fahrenheit, tensile strength measured in pounds per square inch, or any other variables measured on an interval or ratio scale.



PROBLEMS

- 1.1** For each of the following, indicate whether it is a discrete or a continuous variable.
- (a) The number of minutes it takes to read a page in this text.
 - (b) The number of chapters in the text.
 - (c) The weight of the text.
 - (d) The number of problems in the text.
 - (e) The number of times the letter *e* appears on a page.
 - (f) The length of a page in inches.
- 1.2** From weakest to strongest, what is the correct order of ordinal, ratio, nominal, and interval levels of measurement? For each of these levels of measurement, provide two examples of variables that can be used to describe a student.
- 1.3** Suppose that the following information is obtained from Cathy Theo on her application for a home mortgage loan at the North Jersey State Bank. For each of the following responses, indicate whether it is a continuous variable and which type of measurement scale it represents.
- (a) Place of Residence: Morristown, New Jersey
 - (b) Type of Residence: Single family home
 - (c) Date of Birth: August 13, 1966
 - (d) Projected Monthly Payments: \$2,479
 - (e) Occupation: Director of Food Chemistry
 - (f) Employer: Venus Candy Company
 - (g) Number of Years at Job: 10
 - (h) Annual Income: \$140,000
 - (i) Amount of Mortgage Requested: \$220,000
 - (j) Term of Mortgage: 15 years
- 1.4** A team of ornithologists is doing field research by using a mist net to capture migrating birds. They collect the following information:
- (a) Species
 - (b) Weight
 - (c) Wing span
 - (d) Condition, either poor, fair, good, or excellent
 - (e) Band ID number
 - (f) Approximate age
- Indicate whether each of these is an attribute measure or a variables measure.
- 1.5** The management of a company manufacturing gourmet cat food would like to improve its process of packaging its food in plastic containers. Among the measures to be collected are the following:
- (a) weight of the filled container;
 - (b) weight of the unfilled container;
 - (c) whether an individual filled container conforms to specifications;
 - (d) number of containers produced without dents;
 - (e) time needed to fill a batch of containers;
 - (f) number of conforming containers filled per day.
- Indicate whether each of these is an attribute measure or a variables measure.



1.5 OPERATIONAL DEFINITIONS

Operational definitions have long been one of the essential elements of *the scientific method*. The Scientific Method is the basis for all scientific exploration and discovery. Once a problem is identified, hypotheses are empirically tested by

experimentation or observation. Such experimentation and observation require careful collection and analysis of data that will refute or support the hypotheses. It is only recently that operational definitions have been recognized as essential to communication in organizations, whether the organization is manufacturing a product or providing a service.

The purpose of an **operational definition** is to convey meaning that will be precisely the same to different individuals and will remain stable over time.

An operational definition defines a characteristic of an object or event, not just by descriptive adjectives such as red, hard, safe, defective, or sufficiently large, but by the active process or set of operations by which the characteristic is evaluated. Operational definitions are characterized by the following:

1. specific criterion or criteria to be applied to an object or event;
2. a specific test or measurement of the object or event—the particular operations or procedures used to test or measure must be specified and agreed to;
3. a decision as to whether the object or event has met the criterion or criteria—all objects or events that meet the criteria fit the definition, all others do not.

We can illustrate the concept of an operational definition as follows. Suppose that the manager of an office with 20 workers was continually mediating disputes between workers about the level of air conditioning in the office. The workers were constantly adjusting the thermostat up and down and arguing about the comfort level of the work area. The group finally agreed to stop fooling around with the thermostat and to adjust it downward only if it was “too hot.” Each of the 21 people involved knew what “too hot” meant. For each of them, “too hot” was when he or she was personally uncomfortable. Individuals differ in their perception of and sensitivity to heat, however, so there were 21 separate definitions of “too hot.” This group agreement to “adjust the thermostat only when it was too hot” did nothing to solve the problem because they lacked an operational definition. Another meeting was called and the 21 people agreed that it was too hot when the thermometer on the wall on the sunny side of the office registered 80°F. They now had an operational definition on which they could base their actions and the decision as to whether they should turn down the thermostat.

Note that the operational definition of “too hot” contained a specific criterion (80°F), a specific test or measurement of the criterion (temperature to be measured by the thermometer on the wall on the sunny side of the office), and a decision as to whether the temperature met the criterion (all readings of 80°F or higher constituted a “too hot” situation). Now they only had to come to agreement as to which wall the sun was shining on. This was an easier problem to solve, because they agreed that it was unnecessary to measure the amount of light falling on wall surfaces on a daily basis and reached a consensus that the sunny side was the left side of the room before noon and the right side after noon. For an operational definition to be useful, individuals or groups not only must be able to understand and agree upon the criteria and terms used in the definition but also must have the means and ability to evaluate objects and events relative to the criteria.



PROBLEMS

1.6 Provide an operational definition for each of the following:

- (a) a good student;
- (b) a good teacher;
- (c) an outstanding employee;
- (d) a round ball.

1.7 Provide an operational definition for each of the following:

- (a) being on time for a meeting;
- (b) fast service;
- (c) soup that is too hot;
- (d) commuting time to school or work.



1.6 SAMPLING

Most research techniques and many statistical process-control techniques involve the use of sampling. A sample is selected, evaluated, and studied in an effort to gain information about the larger population from which the sample was drawn. In Section 1.1, we learned that a sample is defined as a subset or part of a population. Although, by definition, samples will be smaller than the populations from which they are drawn, samples can be very small or very large. A single student can be considered a sample of all students at a given university, while a very large sample consisting of millions of households is selected to respond to a lengthy questionnaire that is part of the decennial census in the United States.

A sample represents a population, and information obtained from a sample is generalized to be true for the entire population from which it was drawn. The validity or accuracy of generalizations from samples to populations depends on how well a sample represents its population. A well-selected sample can provide information comparable to that obtained by a census.

Advantages of Sampling

Studying a sample instead of a population can have several advantages.

1. **Cost**—Samples can be studied at much lower cost. The smaller number of units or individuals involved in a sample requires less time and money to evaluate. Samples can provide affordable, accurate, and useful information in cases where a census would cost more than the value of the information obtained.
2. **Time**—Samples can be evaluated more quickly than populations. If a decision had to wait for the results of a census, a critical advantage might be missed, or the information might be made obsolete by events or changes that took place while the data were being collected and analyzed.
3. **Accuracy**—Any time data are collected, there is a chance for errors to occur. Errors of measurement, incorrect recording of data, transposition of digits, recording of information in the wrong area of a form, and errors in entering data into a computer can all influence the accuracy of results. In general, the larger the data set, the more opportunity there is for errors to occur. A sample can provide a data set that is small enough to monitor carefully and can permit careful training and supervision of data gatherers and handlers.

4. **Feasibility**—In some research situations, the population of interest is not available for study. A substantial portion of the population might not yet exist or might no longer be available for evaluation. In other cases, evaluation of an item requires its destruction. For example, a manufacturer interested in how much pressure could be applied to a part before it cracked could not perform a census without destroying the entire production run.
5. **Scope of Information**—When evaluating a smaller group, it is sometimes possible to gather more extensive information on each unit evaluated.

The validity and, therefore, the usefulness of information obtained from a sample depends on how closely sample values or statistics estimate population values or parameters. If a sample is to accurately represent a population, it must be unbiased. Biased samples will contain systematic error caused by overrepresentation of some types of individuals or units, while other types are underrepresented. Data obtained from severely biased samples can be worse than no data at all, because they can lead to erroneous conclusions.

Types of Samples

Samples are considered to be either probability or nonprobability samples.

Probability Samples

A **probability sample** is one in which the probability or chance of selecting an element from a population for inclusion in the sample is known.

Knowing the probability of an element's being selected is based on the fact that elements are selected by chance, and knowledge of both the size of the population and the method of selection allows one to calculate the probability of selection. There are many types of probability samples. The simplest probability sample and one of the most widely used is the **simple random sample**. A simple random sample is one in which

1. all elements of a population have an equal chance of being selected for inclusion in the sample;
2. all possible samples of a given size have an equal probability of being selected.

Selection of a simple random sample often involves the use of a table of random numbers or random digits, such as the one provided in Appendix A, Table A.1. Tables of random numbers are constructed so that each of the digits from 0 through 9 has an equal probability of appearing in each position. Tables and lists of random numbers can be generated by computer and are tested statistically to insure that the requirements of independence and equal probability are satisfied. A portion of the larger table in Appendix A is reproduced in Table 1.1 on page 12, to illustrate how such tables can be used to select a simple random sample from lists of the population.

Suppose that a company manufacturing an "ice melt" product that was expected to melt snow and ice at temperatures as low as 15 degrees Fahrenheit was interested in testing the product's ability to perform as expected. A recently manufactured batch of the product consisted of 5,000 ten-pound bags. Here, the advantages of sampling should be self-evident. Not only would it be very time consuming to actually test every bag of the product, but also, because the test involves the use of the product in a destructive manner, if all 5,000 bags were tested, there

TABLE 1.1

USING A TABLE OF RANDOM NUMBERS

ROW	Column		11111	11112	22222	22223	33333	33334
	00000	00001						
	12345	67890	12345	67890	12345	67890	12345	67890
01	49280	88924	35779	00283	81163	07275	89863	02348
02	61870	41657	07468	08612	98083	97349	20775	45091
03	43898	65923	25078	86129	78496	97653	01550	08078
04	62993	93912	30454	84598	56095	20664	12872	64647
05	33850	58555	51438	85507	71865	79488	76783	31708
06	97340	03364	88472	04334	63919	36394	11095	92470
07	70543	29776	10087	10072	55980	64688	68239	20461
08	89382	93809	00796	95945	34101	81277	66090	88872
09	37818	72142	67140	50785	22380	16703	53362	44940
10	60430	22834	14130	96593	23298	56203	92671	15925
11	82975	66158	84731	19436	55790	69229	28661	13675
12	39087	71938	40355	54324	08401	26299	49420	59208
13	55700	24586	93247	32596	11865	63397	44251	43189
14	14756	23997	78643	75912	83832	32768	18928	57070
15	32166	53251	70654	92827	63491	04233	33825	69662
16	23236	73751	31888	81718	06546	83246	47651	04877
17	45794	26926	15130	82455	78305	55058	52551	47182
18	09893	20505	14225	68514	46427	56788	96297	78822
19	54382	74598	91499	14523	68479	27686	46162	83554
20	94750	89923	37089	20048	80336	94598	26940	36858
21	70297	34135	53140	33340	42050	82341	44104	82949
22	85157	47954	32979	26575	57600	40881	12250	73742
23	11100	02340	12860	74697	96644	89439	28707	25815
24	36871	50775	30592	57143	17381	68856	25853	35041
25	23913	48357	63308	16090	51690	54607	72407	55538

Source: Partially extracted from The Rand Corporation, *A Million Random Digits with 100,000 Normal Deviates* (Glencoe, Ill.: The Free Press, 1955).

would be no product remaining to sell. Thus, we can begin by defining a labeling scheme for the 5,000 bags so that each will have a unique identification. The identified bags will serve as our sampling frame. Once the frame is assembled, we can assign a unique number to each of the members of the frame. Because we have 5,000 bags, we will need a four-digit number to identify each member of the frame uniquely. We can use numbers starting at 0001 and ending at 5000. We would then enter the random number table at some random starting point. Closing one's eyes and marking a spot on the table with a pencil is one method that can be used to locate our random starting point. Suppose we mark row 6, column 2. Once we have our starting point, we can move in any direction, up or down columns, across rows, or even diagonally. For this example, we will go across rows. Each member of our frame is identified with a four-digit number, so we will need to select four-digit numbers from the table. The first four-digit number starting at row 6, column 2 is 7340. Whenever we reach a number that is outside the range of the numbers we have assigned to our frame, we disregard it and continue on to the next random number in the table. Because 7340 is out-

side the range of our frame, the highest value of which is 5000, we discard or ignore this number and continue reading across rows until we find a number that is represented in our frame. The next four-digit number, 0336, is in our frame, and bag number 336 is selected for inclusion in our sample. The next number, 4884, is also represented in our frame, so bag number 4884 is the second bag selected. Continuing across the row, we find the next number is outside the range of our frame, so it is disregarded. We continue to row 6, columns 18 to 21 and select bag 3346 for inclusion in the sample. This procedure continues until the number of bags selected is 50, which is the number we had decided is necessary to make up an adequate sample for our purposes.

It is often easier to select numbers by reading down columns in a systematic way. For example, if we randomly select row 16, column 11 as our starting point, we could use the digits in columns 11 to 14 in row 16 and select bag number 3188 for inclusion in the sample. We could then continue down to row 17 and, using the same columns of digits, select bag number 1513. Row 18 provides number 1422. The number provided by row 19 is disregarded because it is outside the scope of the frame. We could continue selecting bags in this manner until 50 have been selected.

Sampling With and Without Replacement

In selecting four-digit numbers from the table, it is possible to obtain a specific number more than once. When this happens, the action we take depends on whether we are sampling with or without replacement. If we are **sampling with replacement**, a member of the population can be represented more than once in a sample. In such cases, any member of the frame whose number occurs twice is selected and is evaluated a second time. If we are **sampling without replacement**, we disregard any numbers that have already been selected for inclusion in the sample.

Sometimes the elements of a frame are prenumbered or have a unique number, such as a serial, invoice, or other identification number assigned for another purpose. As long as such numbers are unique, in that no two elements share the same number and no element has more than one number assigned to it, these numbers can be used as a basis for selection.

Systematic Random Samples

Another type of probability sample that is widely used is a systematic random sample.

A **systematic random sample** is one in which, starting from an element selected at random, every *k*th element in the frame is selected for inclusion in the sample.

If the frame contains *N* elements and the researcher needs a sample of size *n*, *k* can be determined by dividing the size of the frame by the desired sample size (that is, $k = N/n$) and rounding to the nearest integer.

Systematic random samples are often easier to select and can be just as representative of a population as a simple random sample. It would be easier to select every 20th bag in a stack of bags than the 9th, 42d, 91st, 214th, and so on. Similarly, it would be easier to instruct a clerk or computer to select every *k*th unit of output than to use a table of random numbers.



Another way of selecting a systematic random sample is to make k some standard time interval instead of a fixed number of elements. This method is often used when the flow of elements past the sampling and inspection point is rapid; that often occurs in high-volume production processes.

Although a random starting point in the frame is desirable, one is possible only when the frame is such that the researcher can go back to the beginning of the list. Sometimes this is not possible, as when sampling fish passing through a fish ladder on a dam, or when parts reach the end of a line and are ready to be packed and shipped. In such cases, one can start at the beginning of the element flow, provided that the sampling continues long enough to give each element in the frame a chance of being selected.

Systematic random samples can be biased if the elements in the sampling frame are arranged in a pattern characterized by k or if there is a periodicity to the frame or flow of elements past the inspection point. Periodicity in the frame can be a serious problem when k is a time interval instead of a fixed number. For example, if a service organization chose to sample all customers requesting information every seventh day, a biased sample could result, because customers who call on weekends or on Mondays may not be representative of the total population of customers.

Nonprobability Samples

Nonprobability samples are not as scientifically rigorous as probability samples and should be used only when it is impossible to perform a probability sample or when a probability sample would not be cost effective. Among the types of nonprobability samples are convenience samples, judgment samples, and quota samples.

Convenience samples, as their name implies, are based on easy availability for evaluation of members of a population. It is based on the principle of "taking what you can get" quickly and easily. For example, the dean of a school who was interested in student attitudes toward a proposed curriculum change might distribute a short questionnaire to students who happen to be gathered in a lounge area. Such a sample may not be truly representative of all students attending the school, because students who frequent the student lounge may be different from those who do not, but such a sample would be easier to obtain than a random sample of students.

Judgment samples rely on the opinion or judgment of some expert. Members of the population are selected for inclusion in the sample because they are judged to be typical or representative of the population by an individual with some expertise or experience with the population.

When **quota samples** are used, members of the population are selected based on characteristics judged to be important to the study. Individuals are selected so that the proportions falling into each of the categories of the characteristics are equal to the proportions found in the population.

Nonprobability samples can have the advantages of convenience, speed, and lower cost. The disadvantages are lack of accuracy due to bias and limited generalizability of results. Probability samples should be used whenever possible, and nonprobability samples should be used only in situations where one wants rough approximations at low cost or in initial or pilot investigations that will later be followed up by more rigorous research.

PROBLEMS

- 1.8 Suppose that a soft drink bottling company wanted to take a sample of the 20,000 filled bottles that are stored in inventory at a bottling plant. Each bottle is identified by a five-digit identification number and by a code that indicates which of the 20 types of soft drink is contained in the bottle. For each of the following, indicate the type of sample being employed:
- (a) a sample of every fiftieth bottle at the bottling plant;
 - (b) a sample of the first fifty bottles filled on a given day at the bottling plant;
 - (c) a random sample of bottles taken from all 20,000 bottles in inventory.
- 1.9 Given a population with $N = 97$ numbered from 01 to 97, draw a sample of $n = 10$ starting in column 1, row 12 of the table of random numbers (Table A.1) and reading across rows.
- (a) Sample without replacement. List the ten random numbers selected.
 - (b) Sample with replacement. List the ten random numbers selected.
 - (c) Are there any differences between the results in (a) and (b)? Explain your answer.
- 1.10 Given a population with $N = 5,000$, numbered from 0001 to 5000, draw a sample of $n = 15$ starting in column 1, row 5 of the table of random numbers (Table A.1) and reading down columns.
- (a) Sample without replacement. List the 15 random numbers selected.
 - (b) Sample with replacement. List the 15 random numbers selected.
 - (c) Are there any differences between the results in (a) and (b)? Explain your answer.



1.7 STATISTICAL AND SPREADSHEET SOFTWARE

The availability of computer technology has profoundly changed the field of statistics in the last 30 years. Mainframe packages, such as SAS and SPSS, became popular during the 1960s and 1970s. During the 1980s, the field of statistical software experienced a vast technological revolution. In addition to the usual improvements manifested in periodic updates, the availability of personal computers led to the development of many new packages that used a menu-driven interface, and personal-computer versions of existing packages such as SAS, SPSS, and MINITAB quickly became available. The late 1980s and the 1990s represented a continuing period of technological advances. Rapid advances in computer hardware meant that larger amounts of computer memory were available at lower cost. This capacity enabled package developers to include additional, more sophisticated statistical procedures in each subsequent version of their packages. In addition, the advent of desktop productivity tools, such as the Microsoft Excel spreadsheet application, has altered the decision-making processes of many engineers and scientists. Today, an ever-increasing number of engineers and scientists use spreadsheet applications as the means to directly retrieve and analyze data. Because the use of these software packages is now so commonplace throughout the business, academic, and research communities, in this text we will illustrate output from a statistical package, MINITAB, and also from a spreadsheet package, Microsoft Excel. In addition, appendices that follow chapters of this text will contain explanations of how to use MINITAB and Microsoft Excel for the topics that have been discussed in the chapter. Although statistical and spreadsheet software has made even the most sophisticated analyses feasible, problems arise when statistically unsophisticated users who do not understand the assumptions behind procedures or the limitations of results



Another way of selecting a systematic random sample is to make k some standard time interval instead of a fixed number of elements. This method is often used when the flow of elements past the sampling and inspection point is rapid; that often occurs in high-volume production processes.

Although a random starting point in the frame is desirable, one is possible only when the frame is such that the researcher can go back to the beginning of the list. Sometimes this is not possible, as when sampling fish passing through a fish ladder on a dam, or when parts reach the end of a line and are ready to be packed and shipped. In such cases, one can start at the beginning of the element flow, provided that the sampling continues long enough to give each element in the frame a chance of being selected.

Systematic random samples can be biased if the elements in the sampling frame are arranged in a pattern characterized by k or if there is a periodicity to the frame or flow of elements past the inspection point. Periodicity in the frame can be a serious problem when k is a time interval instead of a fixed number. For example, if a service organization chose to sample all customers requesting information every seventh day, a biased sample could result, because customers who call on weekends or on Mondays may not be representative of the total population of customers.

Nonprobability Samples

Nonprobability samples are not as scientifically rigorous as probability samples and should be used only when it is impossible to perform a probability sample or when a probability sample would not be cost effective. Among the types of nonprobability samples are convenience samples, judgment samples, and quota samples.

Convenience samples, as their name implies, are based on easy availability for evaluation of members of a population. It is based on the principle of "taking what you can get" quickly and easily. For example, the dean of a school who was interested in student attitudes toward a proposed curriculum change might distribute a short questionnaire to students who happen to be gathered in a lounge area. Such a sample may not be truly representative of all students attending the school, because students who frequent the student lounge may be different from those who do not, but such a sample would be easier to obtain than a random sample of students.

Judgment samples rely on the opinion or judgment of some expert. Members of the population are selected for inclusion in the sample because they are judged to be typical or representative of the population by an individual with some expertise or experience with the population.

When **quota samples** are used, members of the population are selected based on characteristics judged to be important to the study. Individuals are selected so that the proportions falling into each of the categories of the characteristics are equal to the proportions found in the population.

Nonprobability samples can have the advantages of convenience, speed, and lower cost. The disadvantages are lack of accuracy due to bias and limited generalizability of results. Probability samples should be used whenever possible, and nonprobability samples should be used only in situations where one wants rough approximations at low cost or in initial or pilot investigations that will later be followed up by more rigorous research.

PROBLEMS

- 1.8 Suppose that a soft drink bottling company wanted to take a sample of the 20,000 filled bottles that are stored in inventory at a bottling plant. Each bottle is identified by a five-digit identification number and by a code that indicates which of the 20 types of soft drink is contained in the bottle. For each of the following, indicate the type of sample being employed:
- (a) a sample of every fiftieth bottle at the bottling plant;
 - (b) a sample of the first fifty bottles filled on a given day at the bottling plant;
 - (c) a random sample of bottles taken from all 20,000 bottles in inventory.
- 1.9 Given a population with $N = 97$ numbered from 01 to 97, draw a sample of $n = 10$ starting in column 1, row 12 of the table of random numbers (Table A.1) and reading across rows.
- (a) Sample without replacement. List the ten random numbers selected.
 - (b) Sample with replacement. List the ten random numbers selected.
 - (c) Are there any differences between the results in (a) and (b)? Explain your answer.
- 1.10 Given a population with $N = 5,000$, numbered from 0001 to 5000, draw a sample of $n = 15$ starting in column 1, row 5 of the table of random numbers (Table A.1) and reading down columns.
- (a) Sample without replacement. List the 15 random numbers selected.
 - (b) Sample with replacement. List the 15 random numbers selected.
 - (c) Are there any differences between the results in (a) and (b)? Explain your answer.



1.7 STATISTICAL AND SPREADSHEET SOFTWARE

The availability of computer technology has profoundly changed the field of statistics in the last 30 years. Mainframe packages, such as SAS and SPSS, became popular during the 1960s and 1970s. During the 1980s, the field of statistical software experienced a vast technological revolution. In addition to the usual improvements manifested in periodic updates, the availability of personal computers led to the development of many new packages that used a menu-driven interface, and personal-computer versions of existing packages such as SAS, SPSS, and MINITAB quickly became available. The late 1980s and the 1990s represented a continuing period of technological advances. Rapid advances in computer hardware meant that larger amounts of computer memory were available at lower cost. This capacity enabled package developers to include additional, more sophisticated statistical procedures in each subsequent version of their packages. In addition, the advent of desktop productivity tools, such as the Microsoft Excel spreadsheet application, has altered the decision-making processes of many engineers and scientists. Today, an ever-increasing number of engineers and scientists use spreadsheet applications as the means to directly retrieve and analyze data. Because the use of these software packages is now so commonplace throughout the business, academic, and research communities, in this text we will illustrate output from a statistical package, MINITAB, and also from a spreadsheet package, Microsoft Excel. In addition, appendices that follow chapters of this text will contain explanations of how to use MINITAB and Microsoft Excel for the topics that have been discussed in the chapter. Although statistical and spreadsheet software has made even the most sophisticated analyses feasible, problems arise when statistically unsophisticated users who do not understand the assumptions behind procedures or the limitations of results

are misled by computer-generated statistical output. For pedagogical reasons, we believe that it is important that the applications of the methods covered in the text be illustrated through the use of worked-out examples.



1.8 INTRODUCTION TO QUALITY

As economies and markets have become more interdependent, industrial sectors that have scientific and industrial components have had to look beyond their traditional local competitors and realize that they must compete on a global basis. This need applies to individual manufacturing, research, or service businesses as well as to regional and national economies. Quality and productivity and the relationship between them become essential for survival and growth in this environment. Productivity is fundamental to economic survival, and an increasing body of evidence supports the fact that the key to productivity is quality. In the remainder of this chapter, we will explain both the relationship between quality and productivity and how global competitiveness requires new ways of thinking and new approaches to management and to problem solving. We will also show how statistics has played a historical role in quality and is essential to the understanding and practice of quality-improvement philosophies.

Quality is defined by the individual needs, expectations, perceptions, and experience of the customer. An overall definition of quality is fitness for use.

For a product or service to have value, it must be useful in the sense that it fills a customer's need and that it functions as the customer expects it to function. Customers have traditionally been thought of as the end users or purchasers of a product or service. In a quality environment, it is important to broaden this definition of customers to include all persons within an organization who receive or are affected by the output of your work. Three broad types of quality can be considered: quality of design, quality of conformance, and quality of performance.

Quality of design refers to intentional differences between products or services that are designed by an engineer or planner.

Products or services can have a variety of intentional grades, levels, or styles and come with or without a variety of special or additional features. Differences in quality of design can be based, for example, on the type of material that is used in the manufacture of a product. An inexpensive plastic pen and a gold or gold-plated pen perform the same function and may perform equally well in most circumstances, but they differ in quality of design. The gold pen will be more valuable because of its gold content, attractiveness, durability or style, and it will be considered by many to be of higher quality.

Good quality of design starts with consumer research to assess the needs and expectations of customers and involves close communication between customers, management, designers, and suppliers. The research methods involved in obtaining good quality of design depend on an understanding of statistical methods such as sampling, analysis, summarization, and presentation of data. Communication and feedback from the customer should continue on an ongoing basis in an effort to improve quality and customer satisfaction continually.

Quality of conformance refers to the degree to which a product or service conforms to, meets, and exceeds the standards set forth in the design specifications.

The more consistently and uniformly the standards are met, the higher the quality of conformance. The relationship between uniformity (consistency) and quality may not be as obvious as the relationship between intentional design and quality, but, in many ways, it is more important. A product could be designed to meet functional requirements, have a variety of options, be made of luxurious material, and be the height of current style. Although consumers often prefer to have several choices and options when selecting a product or service, they do not want variability in outcomes once they have made their selection. No matter how excellent in terms of quality of design, if the manufactured products or their component parts are highly variable, even within technical specification limits set out in the design, the product may not function consistently and reliably (or at all). It will be perceived to be (and, in fact, will be) low in quality. The level of quality of conformance is determined by the variability of the manufactured product and its constituent components.

The greater the variability, the lower the quality. To increase quality of conformance, we must decrease variability. Variability can be controlled and reduced if the process by which the product is manufactured or the service provided is continually monitored and analyzed by using statistical methods. Quality of conformance is the responsibility of everyone involved: managers, scientists, engineers, and so on. If statistical monitoring of a process indicates that an unacceptable amount of variability is being produced, the only way to reduce variability is to change the process. Most substantive changes in a process can be made only by management, so reduction of variability is the responsibility of management. The statistical tools that enable managers and engineers to monitor, control, and ultimately reduce variability, thereby improving quality of conformance, will be discussed in Chapters 6 and 7.

Quality of performance refers to the long-term consistent functioning of the product and the related product characteristics of reliability, safety, serviceability, and maintainability.

The degree of quality of performance is closely related to and limited by quality of conformance. It is also affected by the behavior of the manufacturer and the customer after the product has been produced. Failure of the manufacturer to meet responsibilities for such things as delivery, installation, service, parts, and warranty will reduce the quality of performance. Failure of the customer to adequately meet periodic service and maintenance requirements or misuse of the product will also result in a decrease in the quality of performance. If the owner of an automobile with excellent quality of design and quality of conformance fails to regularly change the oil, check the transmission fluid, check the air in the tires, or meet other maintenance requirements set forth by the manufacturer, the automobile will not function optimally and will eventually not function at all. Quality of performance is the responsibility of both management and customers.

It is the responsibility of management to assess quality of performance over time by monitoring such areas as after-sales service calls, maintenance records, and feedback on customer satisfaction. Management's responsibility for quality does not end with the sale of a product or service. Assessment of quality of performance

should continue throughout the life of the product and requires ongoing communication between management and the customer. Survey research techniques and statistical methods are widely used to facilitate this communication and to provide the information that management requires for strategic planning, new-product development, redesign of products and services, and increased quality of performance.



PROBLEMS

1.11 For each of the following, indicate the type or types of quality involved:

- (a) a “luxury” car comes with manual locks and windows;
- (b) a cook bakes a cake using a prepared mix that specifies that the cake should be baked at 350° but sets the oven at 375°;
- (c) An automobile manufacturer produces its new line of cars only in two-door models, while consumer research indicates that families with young children prefer four-door models;
- (d) a case of ball bearings designed to be 10 mm in diameter contains bearings as small as 8 mm and as large as 12 mm in diameter;
- (e) a mail-order birthday gift arrives two days late;
- (f) a child’s bicycle comes unassembled, and the instructions for assembly are unclear or missing;
- (g) a manufacturer produces ski jackets that are the same in terms of basic style but come in a choice of down, wool, or Thinsulate™ insulation.



1.9 A HISTORY OF QUALITY AND PRODUCTIVITY

Although there is no written record, a need for leadership and some notion of quality must have existed among tribal bands of early men and women. The earliest economies of these hunting and gathering groups required ever-increasing communication and coordination between members. Even in a gathering economy, we can imagine quality playing a role, as members, through experience, consider some sticks, stones, or other materials to be better and more fit for use than others. As economies developed from finding tools or products to making them, concepts of quality must have developed and become more important. Once tools that had multiple parts requiring assembly were developed, such as a stone ax head inserted into a wooden handle or a spear point joined to a shaft, the idea of quality must have developed to include the concept of a primitive or rough design standard or specification. Although individuals made their own tools to meet their individual needs, parts had to be made to fit together. If one component broke, wore out, or was lost, a suitable replacement had to be found or made. Although design specifications in terms of quantifiable measurement had not yet developed, rough concepts of standardization, interchangeability of parts, and tolerance limits in terms of fit must have begun to develop.

In craftsman or cottage-industry economies, artisans and farmers who produced products lived in close proximity and communication with their customers and were well aware of their customers’ needs and expectations. Fitness for use and a perception of good workmanship on the part of the artisan defined quality. Although one-of-a-kind products were produced, many everyday utilitarian products made by the same person followed the same design or model and were very similar. This early specialization of individuals who concentrated on making one class of products was an early form of division of labor and increased the productivity of the

overall economy. These artisans and farmers were self-managing and were responsible for providing the quality that their market demanded.

As the capacity to produce goods grew so great that more than local needs could be met, trade developed, inventories were kept, and customers were billed. These activities all required record-keeping. Early examples of the technology of writing are frequently lists and inventories kept for trade or tax purposes. Historical endeavors on a large scale, such as the building of pyramids or temples by the Egyptians, by the Aztecs in Central America, and by the Greeks, and the network of Roman roads all required planning and coordination. Work was supervised by managers and careful measurements were made. These formal measurements are an early example of monitoring work to ensure adherence to specifications and ensure quality.

During the Middle Ages, the growth of guilds occurred, and the apprenticeship system developed. Master craftsmen supervised and directed the training and the quality of work of apprentices and journeymen. Guilds not only managed the training of members in a craft but kept track of the needs of their customers and determined standards of quality and workmanship for the products made. The reputation of the guild rested on the quality of the work of its members. Responsibility for standards and quality rested not just with the individual craftsman—it was also monitored by the guild.

The industrial revolution and the beginnings of the factory system increased the need for standardization of products and separated owner-managers from workers. As early as 1776, Adam Smith in *The Wealth of Nations* (Reference 19) pointed out the increases in productivity and quantity of work that would result from the division of labor. He attributed this increased productivity to the increase in dexterity of the workers performing one task, to the time saved when workers did not have to switch from one task to another, and to the invention and use of machines that facilitated labor. Factories required ever-increasing amounts of raw materials. Monitoring of the quality, consistency, and supply of raw materials became more important. This emphasis increased the need for data gathering and increased the responsibility and role of managers beyond the supervision of workers. Initially, the factory system was limited. Even in industries such as textile manufacture, it coexisted with individual craftspeople. It wasn’t until the late nineteenth and early twentieth centuries that the modern industrial system developed and that manufacturing of products was taken out of the hands of the individual worker or artisan. As the size of an enterprise increased, departmental specialization developed, and management of the enterprise grew beyond the ability of individual owner-managers to control. New levels and methods of management were required.

Frederick W. Taylor tried to address some of the problems created by the factory system and the need to manage large numbers of uneducated and unskilled workers. He is credited with developing the school of Scientific Management in the United States. Scientific Management attempted to increase productivity while benefiting workers, by making it easier for them to perform their job. It required studying the work to be done in detail, determining the optimal methods for performing a job, and carefully training workers to perform each of the steps required. Taylor was concerned that workers were learning their jobs by watching others work and imitating them, rather than by standardized training. This method of learning by watching was traditional and predated the Middle Ages. Inherent in this system was the problem of worker training worker, which is still an issue today. Although effective in a craft or guild economy, workers trained this way can learn

only as much as the worker who trained them, and unfortunately will usually learn all the things the other worker does idiosyncratically or incorrectly. The factory system required large numbers of workers and standardization of work to be effective. Taylor saw that the old methods were no longer effective and that it was management's responsibility to specify the best way to do a job and to train and supervise workers.

Unfortunately, as it developed, Scientific Management had the effect of further separating management from workers. It took responsibility for quality out of the hands of the individual worker and put it into the hands of inspectors, supervisors, foremen, and industrial engineers. The research necessary to analyze a job required a great deal of observing, recording, and analyzing of data. The method of controlling and assuring quality was mass inspection of items produced.

The development of the moving assembly line by Henry Ford further increased the division of labor by breaking down complex jobs into ever more limited and simple tasks that could be performed by unskilled labor. Although this change greatly increased productivity, workers became even less responsible for the quality of the product. Thus, the need for inspection grew. In such circumstances, quality was maintained by weeding out defective or unacceptable items. The underlying or real causes of defects were not sought out and identified. Defects were usually blamed on the worker, and punitive action was taken.

By the 1920s, alternatives to 100-percent inspection began to be considered. At Bell Telephone Laboratories, Walter A. Shewhart, Harold F. Dodge, and Harold G. Romig developed statistically based methods of quality control. Shewhart understood the effects of uncontrolled variability and the losses it caused. In 1924, he introduced the concept of statistical quality control and developed the statistical control chart that, as we will see later in Chapters 6 and 7, became the key to reduction of variability. In 1929, Dodge and Romig developed procedures for statistically based acceptance sampling that eliminated the need for 100-percent inspection.

World War II demanded increased productivity from American technology, industry, and agriculture. Productivity and quality became more important, and statistical methods for the control of quality were widely employed in a variety of industries. Some of the statistical techniques developed during the war were considered so essential to the war effort that they were classified *top secret* by the United States government.

The year 1946 saw the formation of the American Society for Quality Control. The end of the war also found much of the industrial base devastated in Europe and, especially, in Japan. The United States was the only industrial economy not damaged by the war, and, consequently, demand for its products was high. In a competitive free market or consumers' market, quality level tends to be set by the customer; in a producers' market, quality tends to be set by the manufacturer. Many American industries at the time, realizing they could sell just about anything, regardless of quality, abandoned their commitment to quality and the use of the quality-assurance and quality-control methods that had helped win the war. Emphasis shifted from quality to marketing, advertising, and creating demand in order to "push products out the door." The practice was to meet demand with little regard for quality.

In 1950, the Union of Japanese Scientists and Engineers (JUSE) invited W. Edwards Deming to come to Japan to help them rebuild their industrial base and improve the quality image of Japanese products. At the time, the phrase "Made in Japan" had become synonymous with inferior workmanship and poor quality. Deming showed the Japanese that, by proper use of statistical quality-control techniques and a commitment to quality, they would not only recover but

become the best in the world. The work of Deming and of others, such as Armand Feigenbaum, Kaoru Ishikawa, Joseph Juran, and Homer Sarasohn, with their emphasis on quality, is the basis for the rise of Japan to an economic power.

Interest in quality was dormant in the United States until the late 1970s. By that time, many large American companies, such as Motorola, Ford, and Xerox, found themselves under severe competitive pressure from Japanese companies who were able to manufacture higher quality products at lower prices than their American competitors. In 1980, NBC-TV produced a documentary titled "If Japan Can—Why Can't We?" that included an interview with W. Edwards Deming, who was virtually unknown in his home country. This program sparked a reemergence of interest in quality by American companies, first in the manufacturing sector, then in the service industries. To encourage quality improvement in the United States, the Malcolm Baldrige National Quality Award was approved in 1987, to recognize American companies that excel in quality management. This award is based on seven categories: leadership, strategic planning, customer and market focus, information and analysis, human resource development and management, process management, and business results. Among the companies that have won this award are Motorola, Xerox, Federal Express, Cadillac Motor Company, Ritz-Carlton Hotels, AT&T Universal Card Services, and Eastman Chemical Company. Quality improvement is now considered to be an integral part of the strategy of any company that seriously wants to compete in the global economy. In the next section, we will discuss the basic themes of this approach.



1.10 THEMES OF QUALITY MANAGEMENT

Introduction

Quality management views the meeting of customer needs as a primary goal and focuses on quality to increase productivity, reduce costs, increase profit, and compete effectively in a global market. It is based on the work of Deming and of others who have contributed to the quality movement, such as Armand Feigenbaum, Joseph Juran, Homer Sarasohn, Genichi Taguchi, Kaoru Ishikawa, and Myron Tribus, and it integrates the practices of statistical quality control and participative management. The themes displayed in Exhibit 1.1 are fundamental to quality management:

EXHIBIT 1.1 Themes of Quality Management

1. The primary focus is on process improvement.
2. Most of the variation in a process is due to the system and not to the individual.
3. Teamwork is an integral part of a quality-management organization.
4. Customer satisfaction is a primary organizational goal.
5. Organizational transformation needs to occur to implement quality management.
6. Fear must be removed from organizations.
7. Higher quality costs less, not more, but it requires an investment in training.

Theme 1—The Primary Focus Is on Process Improvement

Management's role is no longer viewed as the managing and controlling of people by motivating, disciplining, and rewarding them. Management's role is managing the process. By process, we mean the following.

A process is a sequence of steps that describe an activity from beginning to end.

Managers, engineers, scientists, and workers must work together to improve the process, reduce variability, and improve quality and productivity. This approach focuses on problem prevention and views problems as opportunities to improve the process, not as embarrassments to be covered up.

One well known approach for accomplishing this is the **Shewhart–Deming cycle** presented in Figure 1.2, also known as the Plan-Do-Study-Act cycle or PDSA cycle. It shows the four steps required to improve a process and illustrates that management's efforts are continuous and never-ending.

The first step is *planning*. Planning is based on data collected to determine customer needs and the quality characteristics customers consider important. If data to answer these questions are not available, they must be collected. Previous data on quality of conformance and quality of performance should also be considered. Such tools as cause-and-effect diagrams, flow charts, Pareto charts, histograms, run charts, control charts, and scatter diagrams, which will be presented in later chapters, can facilitate planning. The planning step involves teamwork between managers and individuals from such areas within the organization as research and development, design, engineering, marketing, operations, purchasing, sales, and customer service. It also can be helpful to include other members of the process from outside the organization, such as suppliers and customers. Once the best plan based on the data at hand is formulated, we can proceed to the second step.

Doing involves implementing the plan agreed to in Step 1. During implementation, careful attention must be given to the operational definitions and specifications set forth in the plan. If we are introducing a new product or service or have made major changes in the process, we might want to do a pilot run and implement the plan on a small scale or a trial basis.

Step 3, *studying*, involves evaluating the process and the output of the process. Statistical process-control charts (discussed in Chapters 6 and 7) can be used at this step to monitor quality characteristics and the variability of the process. It is important to monitor the process at many stages. Checking only final products is not sufficient. Variability introduced into the product or service at an early stage will combine with variability from later stages and in the end will be greater than the sum of its parts. Evaluation only at the end of the process will not indicate where in the process variability was caused, nor will it provide information on what caused it and how we might reduce it. Checking must take place over time, at as many stages of the process as possible. Checking and studying the process never end.

In Step 4, *acting*, examination of the data collected and control charts produced in the studying step will tell us whether or not the process is in a state of statistical control, (see Chapter 6) with only chance or common causes of variability operating. If the process is not in a state of statistical control, special causes are operating. We must work to identify these special causes and eliminate them. It is

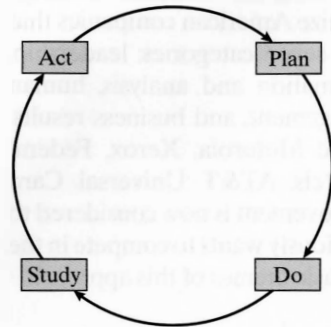


Figure 1.2 The Shewhart–Deming cycle

only when the process is in a state of control that we can consider modifications to the process and attempt to reduce variation and improve quality. In either case, the data will suggest some action to be taken.

From Step 4, we proceed back to Step 1. We now can use what we have learned in previous steps and in the actions suggested by the data to formulate a plan. We continue planning, doing, studying, and acting, indefinitely. The circle that surrounds the steps in Figure 1.2 indicates that the steps are not really separate but flow into one another to form an integrated system of process management. The arrows indicate the direction of the flow, and the fact that there is no break in the circle indicates that the need for process management never ends.

Organizations need to innovate as well as continually improve existing processes. We may have heard how the new technology of the automobile put buggy-whip manufacturers out of business. In fact, not all of them failed. Those that saw the future innovated and, with very little retraining and retooling, made a fortune producing fan belts. Meeting and planning for the future require research on customer needs. Staying in business requires the development of new or higher-quality products and services that will better fill current and future needs. Innovation also requires that resources be put into the maintenance and acquisition of plant and equipment and into a search for new types, sources, and availability of materials.

Theme 2—Most of the Variation in a Process Is Due to the System and Not to the Individual

Both Deming and Juran (References 2, 3, 7, 9, and 23) have pointed out that the causes of most problems and errors (believed to be 85–90%) lie in the system or process and are not under the control of the employee. In such situations, no amount of monitoring of employees' errors and no rewards or punishments based on these counts can improve the system. No employee can do better than the system permits. Analysis of differences in performance ratings in such systems could well say nothing about the true quality of the work of employees or about how skilled or dedicated they are.

Many traditional management theories stress the importance of managers' motivating employees; they assume that motivation is the primary determinant of productivity. If most of the variation is due to the system, however, motivation is of limited value. Workers must know what their jobs are, must have the skills, ability, tools, and resources to do their jobs, and must work with managers and engineers to improve and change the system of operation.

Theme 3—Teamwork Is an Integral Part of a Quality-Management Organization

Division of labor, specialization, and the size of modern enterprises all interacted to fragment organizations into divisions, departments, units, and sub-units. Within these groupings, further separation by functional area and sub-specialization occurred. Work units were separated by floors within buildings, by buildings within a complex, or by geographical distances ranging from across town to hundreds or thousands of miles. Elaborate organizational charts were required just to keep track of what was happening. Very few individuals, if any, understood all the parts or how

they functioned and interacted. The management of each unit, either unaware or unconcerned about the problems of other units, worked to fulfill the goals of its unit, interacting with other units only when necessary—often, that was when problems had occurred.

Under such a system, the work of one area usually got done at the expense of another area. Units naturally (or sometimes by design of top management) were encouraged to compete with one another. Competition was considered to be healthy. Unfortunately, such fragmentation of the process—such physical isolation, competition, and lack of communication—became a barrier between units.

Process management views the satisfying of the customer and the meeting of customer needs as the major goal of an enterprise. *Customers* comprise all the individuals or groups working within the process, as well as vendors, suppliers, and final end users of the product or service that the enterprise provides. When this view is taken, all units and functional areas within the enterprise become customers of one another. They must stay as close to their internal customers as to their external customers. Top management should foster communication and cooperation between staff areas. Managers of staff areas should work together and ensure that their staffs communicate with each other on an ongoing basis.

Teamwork should replace competition at all levels and across all areas. The basic philosophy becomes “none of us is as smart as all of us.” Problem-solving teams whose members come from a variety of management levels and functional areas are an effective means of removing barriers and of encouraging individuals to work together. Managers and their staffs might need training in how to work as teams. Management should see that the resources required for such training are available. Teamwork helps to create trust and remove barriers and creates a sense of long-term commitment to overriding goals.

Theme 4—Customer Satisfaction Is a Primary Organizational Goal

The primary goal of an organization needs to be customer satisfaction. As quality of both manufactured products and services has improved, customers are increasingly unwilling to accept mistakes, poor service, and unreliable products and will simply shift their business to a company that provides reliable products and good service.

Management must be committed to constant improvement in quality of products and services and to improving long-term competitive position. Their goal should be to stay in business, grow, and provide jobs. The long-term benefit must be considered to be not only to the board of directors and the shareholders but to all stakeholders or members of the extended process, such as workers, suppliers, and customers. It is easy to become involved with day-to-day current problems and crises and how to solve them, but, if one is to stay in business, the long-range view and attention to the problems of tomorrow must be considered.

Many organizations that have adopted the philosophy and practice of quality management find that an organizational mission and goals statement that clearly sets out the never-ceasing commitment to constant improvement, reduction of variation, and increased quality insures that all stakeholders understand the constancy of purpose everyone must have in working together for the future.

Theme 5—Organizational Transformation Needs to Occur to Implement Quality Management

The quality-management philosophy is based on continuous quality improvement. Managers and engineers must realize that we have entered a new economic era. If we do not make a commitment to ever-increasing quality and to the increase in productivity that goes with it, we will not be competitive in a global free market. Such changes mean that the organizational culture might need to be altered drastically. Everyone must have the courage to continually examine the way things are done and to be willing to break with tradition and institute new structures and procedures and accept new responsibilities. This is not an easy thing to do, particularly in a period in which a company is enjoying short-term prosperity. It is often the case that it takes a calamity to get organizations and people to change, but it is better to change before the calamity occurs, rather than wait to act while in a weakened position after some negative experiences.

Theme 6—Fear Must Be Removed from Organizations

In the past, management has often used fear as a motivational tool and has ignored, or was unaware of, its economic consequences. There are many kinds of fear that operate in a business environment. There is fear of learning something new. Some managers and engineers are afraid they will not be able to learn new ways. Others are afraid that learning something new means that, all along, they had been wrong. Top management is afraid of looking bad to their board of directors. Middle managers and workers are afraid of speaking out when something is wrong, asking for clarification, or admitting mistakes. These fears help perpetuate the current system and make it impossible for either the company or individual employees to improve. Another type of fear relates to the fear that sharing information may lead to someone else's getting a better rating.

Such fears prevent people from thinking and destroy intrinsic motivation. Often, what will happen in such an environment is that people will not initiate action, because they do not really know what will be viewed positively and what will be viewed negatively. Such fears need to be removed from the workplace in order to achieve quality improvement.

Theme 7—Higher Quality Costs Less, Not More, But It Requires an Investment in Training

Organizations practicing process management must invest resources in education and training. All members of the organization must learn what their jobs are and what is expected of them. They must then be trained to do those jobs. Management's job is managing the process. Although many managers will specialize in one functional area, if they are to be effective they must be familiar with all aspects of the organization and the extended process. At one time, managers worked their way up from entry-level positions within a company. Over the course of their careers, they held a variety of positions in different areas. They knew their organization and how it functioned, and they understood the problems faced by workers at different levels and with a variety of jobs.

Two things occurred to change this practice. One was the increasing specialization within an organization. Division of labor had an effect on managers as well as workers. As companies organized into more specialized functional areas,

managers found that their jobs became narrower. They became specialists in the management of specific limited areas, in a manner similar to the way in which division of labor changed the worker's job from the manufacturing of an entire product to the performing of simple limited tasks. Division of labor took responsibility for the quality of work away from artisans and craftsmen. Specialist management took responsibility for the quality of the entire process away from individual managers.

In addition to managers, workers have a right to know what their jobs are and to receive the training necessary for them to do their jobs well. Their job and the steps necessary to do it should be carefully set forth by using operational definitions. Workers should feel that they are part of an extended process and should understand their role in it. Individuals differ in the style of learning best suited to them and the time necessary for them to learn. Some skills require different methods of training from others. Using experts, management must research the most effective methods to teach a variety of skills to a variety of individuals. Training programs should be innovative and flexible. No one method of teaching or learning will work effectively for all individuals and all jobs.

All members of an organization should be trained in the basic statistical methods required to manage a process. They do not have to become statisticians or experts, but they should be trained to the degree required by their job and their role in the management and improvement of the process. Everyone should understand variability and be able to use basic statistical tools. Organizations that take a long-term view and have a commitment to quality will find that investing in the education and training of all members of the process pays rich dividends.



1.11 THE CONNECTION BETWEEN QUALITY AND STATISTICS

In the preceding three sections of this chapter, we have provided a discussion of the different types of quality, along with the history of quality and the themes of quality management. Awareness and understanding of the statistical concepts and tools of greatest importance to an engineer or scientist will be the subject of the remainder of the text. Familiarity with these statistical tools will enable them to function in organizations that are committed to achieving process improvement and to competing successfully in the global economy.

KEY TERMS

Attribute measures 8	Judgment sample 14
Categorical variables 5	Levels of measurement 5
Census 2	Mutually exclusive 6
Collectively exhaustive 6	Nominal scales 6
Constants 5	Nonprobability samples 11
Continuous variable 5	Numerical variables 5
Convenience sample 14	Operational definitions 9
Data 4	Ordinal scales 6
Data set 4	Parameter 2
Descriptive statistics 1	Population 2
Discrete variable 5	Probability samples 11
Inferential statistics 2	Process 22
Interval scales 6	Quality 16

Quality of conformance 17	Shewhart–Deming cycle 22
Quality of design 16	Simple random sample 11
Quality of performance 17	Statistic 2
Quota sample 14	Statistical thinking 4
Ratio scales 7	Systematic random sample 13
Sample 2	Universe 2
Sampling 10	Variable 4
Sampling with replacement 13	Variables measures 8
Sampling without replacement 13	

CHAPTER REVIEW *Checking Your Understanding*

- 1.12 What is the difference between descriptive and inferential statistics?
- 1.13 How does a population differ from a sample?
- 1.14 What is the difference between a parameter and a statistic?
- 1.15 What is the difference between discrete variables and continuous variables?
- 1.16 What are the various levels of measurement?
- 1.17 What is the difference between attribute measures and variables measures?
- 1.18 What is an operational definition, and why is it so important?
- 1.19 What is the difference between probability and nonprobability sampling?
- 1.20 What are the differences between quality of design, quality of conformance, and quality of performance?
- 1.21 How does quality management differ from more traditional managerial approaches?
- 1.22 Describe the four steps in the Shewhart–Deming cycle.

Problems

- 1.23 For each of the following, indicate which type of measurement scale it represents. Provide an operational definition for each.
 - (a) The distances in feet between the desks in an office.
 - (b) The sizes of the desks in an office.
 - (c) The gender of the workers in an office.
 - (d) The commuting time of the workers in an office.
 - (e) The job classification of the workers in an office.
 - (f) The temperature in the copier room.
- 1.24 For each of the following, indicate which type of measurement scale it represents. Provide an operational definition for each.
 - (a) Weight of an automobile.
 - (b) Exterior color of an automobile.
 - (c) Size classification of an automobile.
 - (d) Original cost of an automobile.
 - (e) Rankings of automobile models by a consumer organization.
 - (f) Trunk size of an automobile.
- 1.25 Suppose that the following information is obtained from each student exiting from the campus bookstore during the first week of classes:
 - (a) Amount of time spent in the bookstore.
 - (b) Gender.
 - (c) Academic major.
 - (d) Class status.
 - (e) Grade-point average.
 - (f) Method of payment.
 - (g) Number of textbooks purchased.

For each of the these, indicate which type of measurement scale it represents. Provide an operational definition for each.

- 1.26 The Data and Story Library (DASL), <http://lib.stat.cmu.edu/DASL>, is an on-line library of data files and stories that illustrate the use of basic statistical methods. Each data set has one or more associated stories. The stories are classified by method and by subject. Access this World Wide Web site and, after reading a story, summarize how statistics has been used in one of the subject areas.
- 1.27 Access the following World Wide Web site provided by Microsoft Corp. for Microsoft Excel (<http://www.microsoft.com/office/excel>). Explain how you think Microsoft Excel could be useful to an engineer or scientist.
- 1.28 Access the World Wide Web site for MINITAB (<http://www.MINITAB.com>). Explain how the use of a statistical software package such as MINITAB might be useful to an engineer or scientist.

WHITNEY GOURMET CAT FOOD COMPANY CASE

Background

It has been estimated by research organizations that there are more than seventy million cats kept as pets in homes in the United States. The market for cat food is an extremely large and lucrative one that has attracted corporations such as Heinz, Nestle, Mars, and Ralston Purina, along with a multitude of small companies that attempt to occupy a market niche with specialty products. Often, these specialty products are sold only in pet-food stores and in veterinarian's offices, not in supermarkets. In recent years, changing lifestyles and attitudes have led to an increasing demand for a higher-quality food product that will be eaten with greater regularity and enthusiasm by the family cat, which is often the only companion for the individual with whom it lives.

Several years ago, the founder of the Whitney Gourmet Cat Food Company, Marilyn Cone, began experimenting with the preparation of home-made cat food for her cat Whitney (for whom the company is named). Her idea was based on the belief that alternate types of meat and seafood, along with a chunkier formulation with less fillers, would lead to higher level of satisfaction (and less finicky behavior) by cats, without the sacrifice of nutritional standards. Within a short period of time, her cottage industry was ready to become commercial, and a processing plant was obtained and developed for producing canned cat food. Currently, two types of food are made, one based on kidney parts, the other formulated with pieces of shrimp. In the five years that the company has been operating commercially, sales have risen rapidly, particularly in the past three years, during which the number of sales outlets has been greatly expanded. The rapid expansion of the company has, however, led to the realization that it was too large to be managed as a cottage industry any longer. Senior management of the company, under the leadership of Ms. Cone, is increasingly conscious of the need to

increase the efficiency of operations and strive for ever-improving quality levels.

Phase 1

A task force consisting of corporate-level officers, engineers, and department heads was formed to consider how to go about the quality-improvement effort. There was agreement that the first step was the development of a mission statement for the company that could succinctly communicate its mission to both customers and employees.

Once the mission statement was developed, the task force turned to a discussion of which areas of operations should be examined for improvement opportunities. After much brainstorming and discussion, the task force decided that one critical area for improvement was the number of cans that were deemed to be nonconforming (for any one of a variety of reasons).

As Al Brenner, the production manager, and a team of other managers and engineers brainstormed about how to go about the quality-improvement process, the group realized that, in the rapid expansion that had taken place in the past three years, perhaps not enough attention had been paid to training workers properly for their jobs. The company had been running two shifts of workers for the past year, in an effort to keep up with demand for the product, and it had several hundred people working on various aspects of the production and canning process. One of the aspects that concerned the group was whether the three inspectors who worked on each shift were applying the current definition for what was to be considered a nonconforming can of cat food consistently. The team decided that an evaluation of the final process, inspection of cans, needed to be carried out.

TABLE WG1.1

INSPECTION TEST RESULTS OF TWELVE CANS OF CAT FOOD

CLASSIFICATION	Inspector					
	Shift 1			Shift 2		
	1	2	3	4	5	6
Nonconforming	4	4	4	4	4	4
Conforming	8	8	8	8	8	8

Al Brenner, the production manager, selected twelve cans that ranged in his opinion from clearly conforming, to marginal, to nonconforming for one of several reasons. The cans were tagged with an identification number and placed on tables set up in a conference room. The six inspectors, three from each of the two shifts, independently were instructed to evaluate the cans in the same way that they did on their job every day. Inspection test results are summarized in Table WG1.1.

Exercises

- WG1.1 What conclusions about the consistency of the inspectors can you draw?
- WG1.2 What other information concerning the inspection test results would be useful to obtain?

STOP

Do Not Read Below until the Phase-1 Exercises have been completed.

Phase 2

The team realized that, in order to study the consistency among the inspectors completely, a tally of the results for each individual can by each inspector had to be available. Table WG1.2 represents a summary of the inspection decisions made by each of the six inspectors for the twelve cans.

Exercises

- WG1.3 What conclusions about the inspection process can you now reach after examining Table WG1.2?
- WG1.4 Set up a table that would help you to analyze the disagreements between the inspectors.

TABLE WG1.2

INSPECTION TEST RESULTS OF 12 CANS OF CAT FOOD, TALLIED BY INSPECTOR

Can	Inspector	Conforming	Nonconforming	Reason
1	1	×		
	2	×		
	3	×		
	4		×	Blemish on top of can
	5		×	Blemish on top of can
	6	×		
2 3*	Accepted by all			
	1		×	Pull tab missing
	2		×	Pull tab missing
	3		×	Pull tab missing
	4		×	Pull tab missing
	5		×	Pull tab missing
	6		×	Pull tab missing

(Continued)

TABLE WG 1.2 (Continued)

INSPECTION TEST RESULTS OF 12 CANS OF CAT FOOD,
TALLIED BY INSPECTOR

Can	Inspector	Conforming	Nonconforming	Reason
4*	Accepted by all			
5*	1	×		
	2		×	Crack in top of can
	3		×	Crack in top of can
	4	×		
	5	×		
	6		×	Improper pull-tab location
6*	1		×	Surface defect on side of can
	2	×		
	3		×	Surface defect on side of can
	4		×	Surface defect on side of can
	5	×		
	6	×		
7	Accepted by all			
8	Accepted by all			
9*	1		×	Crack in side of can
	2		×	Crack in side of can
	3	×		
	4	×		
	5	×		
	6		×	Crack in side of can
10	1	×		
	2		×	Improper location of pull tab
	3	×		
	4	×		
	5		×	Surface defect on side of can
	6	×		
11*	1		×	Crack in top of can
	2	×		
	3		×	Crack in top of can
	4		×	Blemish on top of can
	5		×	Crack in top of can
	6		×	Crack in top of can
12	Accepted by all			

*Classified as nonconforming by production manager

REFERENCES

1. Cochran, W. G.; *Sampling Techniques*, 3d ed.; New York: Wiley, 1977
2. Deming, W. E.; *Out of the Crisis*; Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1986
3. Deming, W. E.; *The New Economics for Industry, Government, and Education*; Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1993

4. Dobyns, L. and C. Crawford-Mason; *Quality or Else*; PBS Video Series, Silver Spring, MD: CC-M Productions, 1991
5. Dobyns, L. and C. Crawford-Mason; *Quality or Else*; Boston, MA: Houghton-Mifflin, 1991
6. Feigenbaum, A. V.; *Total Quality Control*, 3d ed.; New York: McGraw-Hill, 1983
7. Gabor, A.; *The Man Who Discovered Quality*; New York: Time Books, 1990
8. Halberstam, D.; *The Reckoning*; New York: Avon Books, 1986
9. Juran, J. M.; *Juran on Leadership for Quality*; New York: The Free Press, 1989
10. Kendall, M. G. and R. L. Plackett, eds.; *Studies in the History of Statistics and Probability, vol. II*; London: Charles W. Griffin, 1977
11. *Microsoft Excel 2000 for Windows*; Redmond, WA: Microsoft Corp., 1999
12. *MINITAB Version 12*; State College, PA: MINITAB, Inc., 1998
13. Pearson, E. S., ed.; *The History of Statistics in the Seventeenth and Eighteenth Centuries*; New York: Macmillan, 1978
14. Pearson, E. S. and M. G. Kendall, eds.; *Studies in the History of Statistics and Probability*; Darien, CT: Hafner, 1970
15. Rand Corporation; *A Million Random Digits with 100,000 Normal Deviates*; New York: Free Press, 1955
16. Shewhart, W. A.; *Economic Control of Quality of Manufactured Product*; New York: Van Nostrand-Reinhard, 1931, reprinted by the American Society for Quality Control, Milwaukee, WI, 1980
17. Shewhart, W. A.; *Statistical Methods from the Viewpoint of Quality Control*, edited by W. E. Deming; Washington, D.C.: Graduate School, Department of Agriculture, 1939; Dover Press, 1986
18. Skrebec, Q. R.; "Ancient process control and its modern implications," *Quality Progress*, 25 (1990), 49-52
19. Smith, A.; *The Wealth of Nations* (1776); New York: Modern Library, 1937
20. Snee, R. D.; "Statistical thinking and its contribution to total quality," *American Statistician*, 44 (1990), 116-121
21. Taylor, F. W.; "The Principles of Scientific Management," *Bulletin of the Taylor Society*, (December 1916); Reprinted in *Classics of Organizational Theory*, J. Shafritz and P. Whitlock, eds.; New York: Moore Publishing Co., 1978
22. Walton, M.; *Deming Management at Work*; New York: G. P. Putnam and Sons, 1990
23. Walton, M.; *Deming Management Method*; New York: Perigee Books, Putnam Publishing Group, 1986

APPENDIX 1.1 Basics of the Windows User Interface

In this appendix, we first provide an orientation to the basic concepts necessary to operate any program, such as Microsoft Excel or MINITAB, that runs in a Windows user interface in which windows or frames are used as containers to subdivide the screen. In this user interface, although communication can be done with some combination of keystrokes, many pointing or choosing tasks are more easily done by using a pointer device such as a mouse, trackball, or touchpad. Moving a pointer device moves a **mouse pointer**, an on-screen graphic that in its most common form takes the shape of an arrow. Moving a mouse pointer over another object and pressing one of the buttons on the pointer device defines a mouse operation. Four types of mouse operations used in Microsoft Excel or MINITAB are defined in Exhibit A1.1.1 on page 32.

In a windowing environment, mouse operations are applied to a variety of on-screen objects. First and foremost among these objects are **windows**, frames that serve as containers for other windows or for the objects described in this appendix. Many windows can be selected and dragged from one position on the screen to another, and

EXHIBIT A1.1.1 Types of Mouse Operations

1. To **select** an on screen object, move the mouse pointer (by moving the mouse) directly over an object and then press the left mouse button (or click the single button, if your mouse contains only one button). If the on-screen object being selected is a button, **click** is often used as an alternative to the verb *select*, as in the phrase “click the OK button.”
2. To **drag** or move an object, first move the mouse pointer over an object and then, while holding down the left (or single) mouse button, move the mouse. After the object has been dragged, release the mouse button.
3. To **double-click** an object, move the mouse pointer directly over an object and press the left (or single) mouse button twice in rapid succession.
4. To **right-click** an object, move the mouse pointer directly over an object and press the right mouse button. If your mouse contains only one mouse button, right-clicking is the same as simultaneously holding down the control key and pressing the mouse button.

sometimes they are capable of being **resized** (having their length and width dimensions changed).

Other common objects in a windowing environment are free-floating icons, menu bars, task or tool bars, and dialog boxes.

Free-Floating Icons

Free-floating **icons**, graphics that represent a specific application or document, can be selected and dragged from one position to another. Activating an icon representing Microsoft Excel would be a typical way of starting the Excel program.

Menu Bars

Menu bars are the horizontal lists of words that represent a set of choices. Selection of one of the choices results in the display of a **pull-down menu**, a list of more word choices.

Task or Tool Bars

Task or **tool bars** are groups of fixed-position icons or **buttons**, clickable graphics that simulate the operation of a mechanical push button. Tool bars can be either free-floating or fixed (“snapped” into an on-screen position). In applications such as Microsoft Excel, there are tool bars that represent different categories of user actions, such as formatting, editing, or drawing.

Dialog Boxes

Making a certain selection on a menu or tool bar often results in a special type of window called a **dialog box**. Dialog boxes are used to display status messages or to prompt you to make choices or supply additional information. Common to many dialog boxes are the objects listed in Exhibit A1.1.2.

EXHIBIT A1.1.2 Objects Common to Dialog Boxes

- **Drop-down list** boxes allow selection from a nonscrollable list that appears when the drop-down button, located on the right edge of the box, is clicked.
- **Scrollable list** boxes display a list of items, in this case, files or folders, for selection. When the list is too long to be seen in its entirety in the box, clicking the right scroll button or dragging the slider on the scroll bar at the bottom of the list box reveals the rest of the choices.
- **Edit** boxes provide an area into which a value can be edited or typed. Edit boxes are often combined with either a scroll-down list box or spinner buttons, to provide an alternative to the typing of a value.
- **Option** buttons represent a set of mutually exclusive choices. Selecting an option button (also known as a radio button) always deselects, or clears, the other option buttons in the set, thereby allowing only one choice at a time.
- **Check** boxes allow the selection of optional actions. Unlike option buttons, more than one check box in a set can be selected at a given time.
- Clicking the **Open** or **OK** buttons causes an application to execute an operation with the current values and choices as shown in the dialog box.
- Clicking the **Cancel** button closes a dialog box and cancels the operation.

COMMENT: ABBREVIATED NOTATION FOR MENU SELECTION

In this text, the authors will abbreviate menu selections by using the vertical slash character | to separate menu choices. For example, **File | Open** will be used instead of the longer “select the File menu and then select the Open choice.” In addition, as was done with **File | Open**, the actual selections will be in **bold-faced** type.

APPENDIX 1.2 Introduction to Microsoft Excel

Microsoft Excel and This Text

Microsoft Excel is an example of a spreadsheet application program, the type of personal productivity program best suited for the interactive manipulation of numeric data. Spreadsheet applications allow users to make entries onto electronic versions of paper **worksheets**, rectangular arrays of (horizontal) rows and (vertical) columns into which entries are made. They were first used in business, but today the flexibility of modern spreadsheets makes them an everyday problem-solving tool for many, including students learning statistical problem-solving in a course for engineers and scientists.

Of the many spreadsheet programs available, the authors have chosen to use Microsoft Excel in this text, and not just for the obvious reasons that the program is widely available and incorporates the commonly used Microsoft Office user interface. Microsoft Excel also contains special statistical functions and procedures that aid in the analysis of data and can accept add-ins, preprogrammed procedures that extend the functionality of Excel. These features help construct statistical solutions in Excel and simplify its use. (On many occasions, the Excel appendices use the Prentice Hall PHStat add-in supplied on the CD-ROM that accompanies this text, to further streamline the use of Excel, as explained in the Add-ins section of this appendix.)

Microsoft Excel also allows users to create workbooks, collections of electronic worksheets and other information, which are combined into a single disk file. Worksheets facilitate the development of solutions that are consistent with the rules of good application design. In this text, the solutions generated follow the predictable pattern of placing the problem's data, calculations, and graphical objects on separate sheets. These separate-sheet designs, in turn, enhance the reusability of the workbooks, provide easier opportunities for modifying the workbooks, and generally aid in the clarity of the presentation of the results.

Although it is useful as a tool for learning statistics, readers should be aware that using Microsoft Excel is not an all-purpose substitute for using a standard statistical package. Very large data sets, or data sets with unusual statistical properties, can cause Excel, as well as any add-in code being used, to produce invalid results.

Getting Familiar with the Microsoft Excel Application Window

When Microsoft Excel program or workbook icons are double-clicked, Windows loads Excel, and an Excel application window similar to the one shown in Figure A1.2.1 appears.

Users can configure the exact combination and placement of objects that appear in this window. In Figure A1.2.1, as in all illustrations of the Excel application in this text, the standard and formatting toolbars have been configured to appear below the menu bar. The worksheet displayed in the workspace area has been maximized (opened to cover the entire workspace area). Sets of resizing and closing buttons appear on the title bar and the menu bar. Scroll bars, both horizontal and vertical, allow for the display of parts of the worksheet currently off-screen. Sheet tabs identifying the names of the individual sheets provide a means of "turning" to another sheet in the workbook. A status bar displays information about the current operation and the state of certain keyboard toggles.

Specifying Worksheet Locations

Part of the design process when using a spreadsheet application involves placing the data, calculations, and results required, along with such titles and labels as column headings, into a grid formed by individual worksheets. Using a standard notation for worksheets, we refer to the columns of a worksheet by letters and to the rows by numbers, to identify into which cells (intersections of the columns and rows) entries should be placed. Under this system, the cell reference A1 refers to the cell in the first column and the first row (the upper-left-corner cell); cell reference B4 refers to the cell in the second column and the fourth row.

Because a Microsoft Excel workbook can contain multiple worksheets, this column letter and row number format is, in certain contexts, insufficient to specify the particular cell of interest. When a worksheet design calls for cell entries on one worksheet to refer to cells on another worksheet, the cell reference must be written in the form:

Sheetname!ColumnRow

Using this notation, one can distinguish between two similarly located cells of two different sheets in the workbook. For example, Data!A1 and Calculations!A1 refer to the upper-

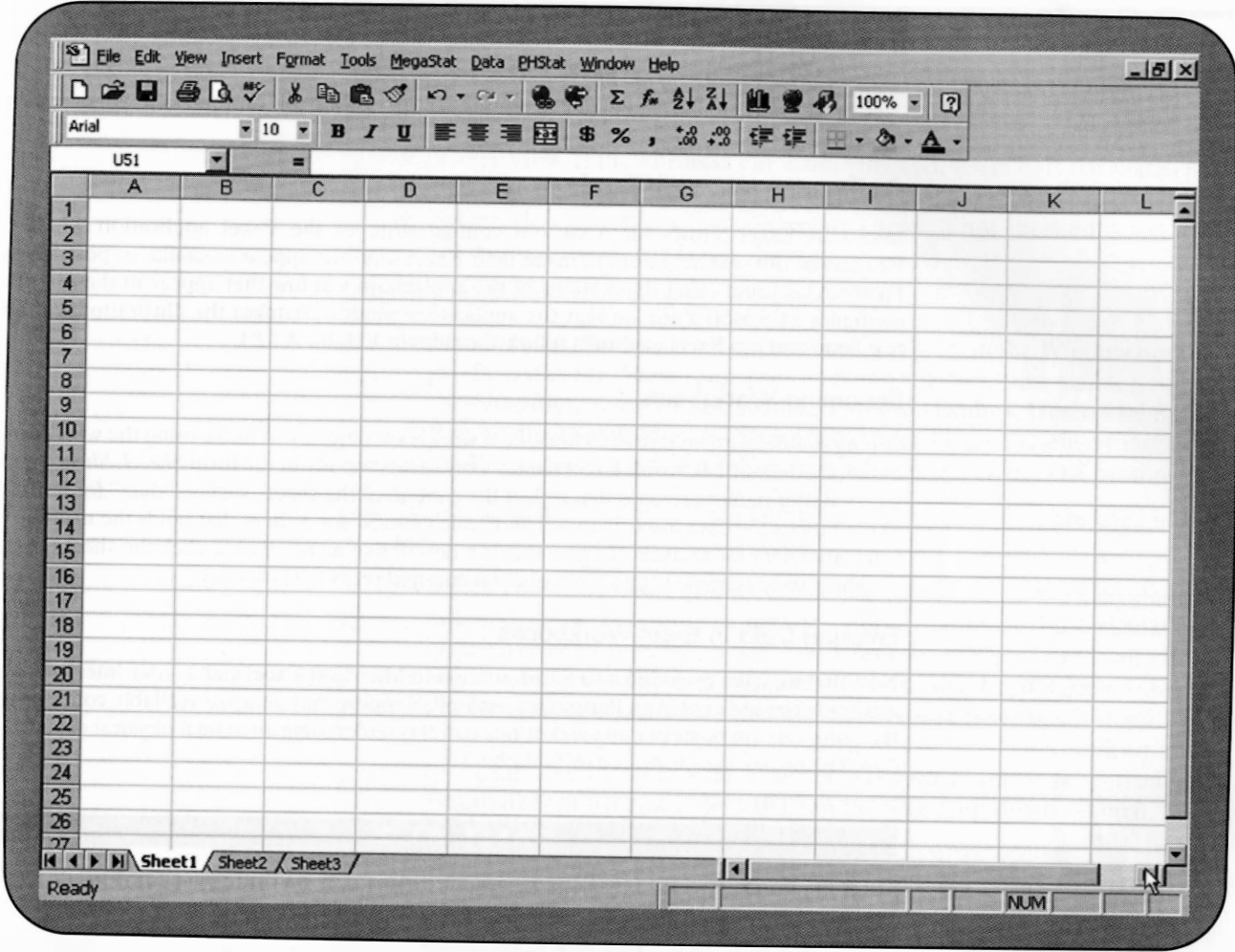


Figure A1.2.1 The Excel application window

left corner cell of the data and calculations sheets, respectively. This extended notation is necessary only when the cell reference is to a worksheet other than the current one, into which entries are being made.

Beside individual cells, references can be made to cell ranges that are rectangular groups of adjacent cells. Cell ranges are written using the cell references to the upper-left-most and lower-right-most cells in the block. The cell range form is

UpperLeft:LowerRight

For example, the cell range A1:B3 refers to the six-cell worksheet block containing the cells A1, B1, A2, B2, A3, and B3; the range A1:A8 refers to the first eight cells in the first column of the worksheet. Ranges in the form

Sheetname! UpperLeft:LowerRight

are allowed and refer to ranges not on the current sheet.

Sometimes, there is a need to distinguish between two similarly located cells on two similarly named worksheets in two different workbooks located in the same folder or

directory. In such cases, cell references are written in the form

`'[Workbookname]Sheetname'!ColumnRow`

—as in `'[FISH]Data'!A1`, referring to the upper-left-corner cell on the Data worksheet in the FISH workbook.

Configuring Microsoft Excel

Microsoft Excel allows for a custom configuration of the Excel application window. Readers of this text may want to make their Excel windows appear as similar as possible to Figure A1.2.1 and other illustrations of the application window that appear in the text. To configure Microsoft Excel so that the application window matches the illustrations in the text, load and run Excel, and then follow the steps in Exhibit A1.2.1.

Renaming Worksheets

Giving self-descriptive names to individual workbook objects can make using the workbook easier. By default, Microsoft Excel names worksheets serially in the form *Sheet1*, *Sheet2*, and so on. Better names are ones that reflect the content of the sheets, such as “data” for a sheet that contains the data to be analyzed and “calculations” for a sheet that holds the necessary calculations for an analysis. To give sheets a descriptive name, double-click the sheet tab of the sheet to be renamed, type a new sheet name, and press the enter key.

Entering Data in Excel Workbooks

Now that we have provided a basic orientation to Microsoft Excel and its user interface, we are ready to use Excel in an illustrative example. Suppose that data are available concerning the pulse rate (in beats per minute) of five patients undergoing an annual physical examination. The results are illustrated in Table A1.2.1.

TABLE A1.2.1	
PULSE RATE (IN BEATS PER MINUTE) OF FIVE PATIENTS UNDERGOING AN ANNUAL PHYSICAL EXAMINATION	
Name	Pulse Rate (in beats per minute)
Allen	62
Barry	70
Diane	72
Kim	58
Susan	52

We are now ready to use Microsoft Excel to perform statistical analysis on these data. For our data of Table A1.2.1, we decide that the first column heading (Name) should appear in cell A1. Using the same reasoning, we determine that the Pulse Rate heading should appear in cell B1 and that the values of each variable for the five patients should appear in the next five rows (2 through 6). Having specified cell addresses for the various parts of the worksheet, we are now ready to enter values into the cells of a sheet that we will name Data. To do this, activate the Microsoft Excel application, and select **File | New** to create a new, blank worksheet window. Rename Sheet1 as **Data**.

Select cell **A1** by clicking its interior. A special border, the **cell highlight**, appears around the cell. This highlight indicates that cell A1 is now the **active cell**, the cell into which the next value to be typed will be entered. (Also note that A1, the address of the active cell, appears in the cell reference box.) Type the column heading Name. As you type, notice that your keystrokes appear both in the edit box of the formula bar and in cell A1 itself. Press

EXHIBIT A1.2.1 Configuring Microsoft Excel

1. To display the formula and status bars and the standard and formatting toolbars:
 - Select **View**. If the **Formula** bar is not checked, select it. If the **Status** bar is not checked, select it.
 - Select **View | Toolbars**. If the **Standard** choice is not checked, select it. If the **Formatting** choice is not checked, select it.

To standardize the display of the worksheet area:

- Select **Tools | Options**. In the Options dialog box, in the Window options group select the **View** tab (if another tab's options are visible). Select the **Gridlines**, **Zero Values**, **Row & Column Headers**, **Horizontal Scroll Bar**, **Vertical Scroll Bar**, and **Sheet Tabs** check boxes. Deselect (uncheck) the **Formulas** check box if it has been selected. Click the **OK** button.
2. To verify calculation, edit, and general display options that are used in the text:
 - Select **Tools | Options**. In the Options dialog box, select the **Calculations** tab, and verify that the **Automatic** option button of the Calculations group has been selected. Select the **Edit** tab, and verify that all check boxes except the Fixed Decimal check box have been selected. Select the **General** tab. Verify that the **R1C1 Reference Style** box is deselected (unchecked). Change the value in the **Sheets in New Workbook** edit box to 3 (if it is some other value). Select **Arial** (or a similar font) from the **Standard Font** list box. Select **10** from the **Size** drop-down list box.

3. To verify the installation of the Data Analysis ToolPak add-in:
 - Select **Tools**. Select **Data Analysis** if it appears on the Tools menu choice, and verify that the dialog box that appears contains an Analysis Tools scrollable list box. Click the **Cancel** button. This verifies the installation of the tools.
 - If Data Analysis does not appear on the Tools menu, select **Add-Ins** from the menu. In the Add-Ins dialog box, select the **Analysis ToolPak** and the **Analysis ToolPak-VBA** check boxes—if these check boxes appear—from the **Add-Ins Available:** list. Click the **OK** button. Exit Excel (Select **File | Exit**). Rerun Excel, and follow the instructions in the previous paragraph to verify installation.
 - If you cannot find the Analysis ToolPak or the Analysis ToolPak-VBA check boxes in the Add-Ins Available: list, most likely these special add-in files were not selected during the Microsoft Excel setup process. For example, if you used the “Typical “ option of the Microsoft Excel/Office setup program, the (Data) Analysis Toolpak add-ins used in this text were not selected, and you would need to run the Microsoft Excel/Office setup program a second time to include them. (See the “install and use the Analysis ToolPak” on-line help topic for further information.)

the **Enter** key (or click the check button to the left of the edit box) to complete the entry. (Users of keyboards that do not contain an Enter key should press the Return key when instructions in this text call for pressing the Enter key.) Continue by selecting cell B1 and entering the column heading Pulse Rate.

Now that the column headings have been entered, we can begin to enter the values that will appear underneath them. We will type in values by columns, using the feature of the Enter (Return) key to automatically advance the cell highlight down one row after each entry. (If we wished to enter values by rows, we would end each entry by pressing the Tab key, which would advance the cell highlight one column to the right.)

Select cell A2, type the name Allen, and press the Enter (Return) key. Then type the remainder of the names in cells A3–A6, being sure to press the Enter key after each name. Select cell B2, and enter the pulse rate of 62 from Table A1.2.1 into cell B2. Then continue by entering the corresponding amounts of 70, 72, 58, and 52 into cells B3 through B6. Having now entered all the data from our source table into the Data worksheet, we should save a copy of our work on disk, by using **File | Save** or **File | Save As**, before continuing.

COMMENT: CORRECTING ERRORS

As you make entries into a worksheet, at some point you will probably make a typing error. To correct typing errors, you can do one of the following:

- To cancel the current entry while typing it, press the **Escape** key, or click the “X” button on the formula bar.
- To erase characters to the left of the cursor one character at a time, press the **Backspace** key.
- To erase characters to the right of the cursor one character at a time, press the **Delete** key.
- To replace an in-text error: first, click at the start of the error; next, drag the mouse pointer over the rest of the error; finally, type the replacement text. If you change your mind, you can undo your last edit by selecting the command **Edit | Undo**. If you change your mind again and wish to keep the edit after all, select **Edit | Redo**.

Developing Formulas to Perform Calculations

Now that we have saved our work, we are ready to compute a simple statistic, the total pulse rate of the five patients. One way to generate the total pulse rate value would be to just manually sum the values, adding 62, 70, 72, 58, and 52 to get 314 as the total pulse rate.

Although it might be argued that for this very small and very simple problem, manual calculation would be the best method to obtain the total amount, more generally it is wiser to have Microsoft Excel generate the calculation than to do it yourself. To have Excel generate these calculations, we will need to develop and enter formulas, or instructions to perform a calculation or some other task, in the appropriate cells of our Data worksheet (cell B7 in this example).

To distinguish them from other types of cell entries, all formulas always begin with the = (equal sign) symbol. Creating formulas requires knowledge of the **operators**, or special symbols, used to express arithmetic operations. Operators used in formulas in this text include addition (+), subtraction (–), multiplication (*), division (/), and exponentiation [a number raised to a power] (^).

Because our definition of the total of the pulse rates calls for the addition of five quantities, we will use the + (plus sign) in our *total pulse rate* formula, combining it with the cell addresses containing the values we wish to add. In the case of calculating the total pulse rate for the five patients, cells B2, B3, B4, B5, and B6 on the Data sheet are added together.

By assembling these pieces, we could form the formula

=Data!B2 + Data!B3 + Data!B4 + Data!B5 + Data!B6

and enter it into cell B7. Because we are entering a formula in the same sheet as the sheet to which it refers, we can instead write the formula by using the shorthand notation

=B2 + B3 + B4 + B5 + B6

and Microsoft Excel will correctly interpret the addresses as referring to the current (Data) sheet.

Using Functions in Formulas

In our discussion of formulas, we used the plus sign arithmetic operator to construct our formula. We could just as easily have used the sum function, one of many such preprogrammed instructions that can be used when solving a variety of common arithmetic, business, engineering, and statistical problems.

To use the Sum function, we type the formula =SUM(B2:B6) into cell B7 instead of the formula =B2 + B3 + B4 + B5 + B6. In the formula =SUM(B2:B6), the word SUM identifies the sum function, the pair of parentheses () bracket the cells of interest, and B2:B6 is the address of the cell range of interest, the cells whose values will be used by the function.

COMMENT: COPYING OBJECTS IN MICROSOFT EXCEL

Objects ranging from a single cell to an entire worksheet can be copied, to simplify or speed implementation of a worksheet design. Generally, copying involves first selecting the object and then selecting the appropriate copy and paste commands.

- To copy a single cell entry or a range of entries, select the cell containing the entries to be copied (or the cell range) by dragging the mouse pointer through all the cells of the range. Select **Edit | Copy**. Select the cell (or select the first cell of the range) to receive the copy. Select **Edit | Paste**. (Note: Copying entries that contain formulas might not result in duplicate entries. See Appendix 3.1 for an explanation of absolute addresses.)
- To copy an entire worksheet, select the worksheet to be copied by clicking on its sheet tab. Select **Edit | Move or Copy Sheet**. In the Move or Copy dialog box, select the **Create a Copy** check box. Select **New Book** from the To book: drop-down list box, if the copy of the worksheet is to be placed in a new workbook. Select the workbook position for the copy by selecting the appropriate choice from the Before sheet: list box. Click the **OK** button.

Wizards

Wizards are sets of linked dialog boxes that guide the user through the task of creating certain workbook objects. Users enter information and make selections in the linked boxes and advance through the set by clicking a Next button (and, ultimately, a Finish button to create the object—clicking a Cancel button cancels the task.) As an example, consider the Microsoft Excel **Text Import Wizard** that assists in the importing, or transferring, of data from a text file into a worksheet. (A text file contains unlabeled and unformatted values that are separated by *delimiters*, such as spaces, commas, or tab characters.).

To demonstrate this wizard, assume that a set of fish weights have been stored in the text file FISH.TXT—such a file is supplied on the CD-ROM that accompanies this book. To import the data from this file into an Excel worksheet, select **File | Open**. In the Open dialog box, select the folder that contains the FISH.TXT file from the Look in: drop-down list. Select the **Text Files (*.prn; *.txt; *.csv)** choice from the Files of type: drop-down list. The All Files (*.*) choice can also be selected. Enter the name **FISH.TXT** in the File name: edit box, or select it from the file list box. (If the file does not appear in the file list box, verify that the previous two steps were done correctly.) Click the **OK** button.

This sequence launches the three-step *Text Import Wizard*. In the Text Import Wizard Step 1 dialog box, select the **Delimited** option button (because the data values for the variables in this file have been placed in fixed-width columns). Note that what this book calls variables are called fields in the dialog box. Click the **Next** button. In the Text Import Wizard Step 2 dialog box, in the Delimiters group, select the space edit box if the variables have been separated with spaces or the Tab edit box if the variables have been separated with a Tab. Click the **Next** button to accept the placement of the data from each line in the text file into columns. (Dragging the vertical column separator line would alter the placement, if necessary.) In the Text Import Wizard Step 3 dialog box, select the **General** option button under the Column data format heading. Click the **Finish** button. The data of the text file is transferred to a new worksheet, named after the text file name (FISH in this example), in a new workbook. Saving the data as a Microsoft Excel workbook by selecting **File | Save As**, save as type **Microsoft Excel Workbook**. Click the **Save** button.

Objects created by wizards should be reviewed for errors. Objects containing errors can be either edited (typically by right-clicking the object and selecting the appropriate choice from the shortcut menu) or else deleted and then recreated by the wizard. (In this example, a good procedure would be to review the data just transferred and, perhaps, insert column headings, too, by first selecting any cell in row 1, then selecting **Insert | Rows**, and finally entering the actual column headings.)

Add-Ins

Add-ins are optional preprogrammed procedures that extend the functionality of Microsoft Excel. Some add-ins, such as the Data Analysis ToolPak previously mentioned, are supplied with the Microsoft Excel program files and can be installed (made permanently available to users) by running the Microsoft Excel setup program. Other, so-called “third party” add-ins, such as the Prentice Hall PHStat add-in that is supplied on the CD-ROM that accompanies this text, are loaded or installed separately.

Add-ins typically modify the menu bar of Microsoft Excel, either by inserting a new menu choice into a pre-existing menu or by inserting a new menu of choices. Once selected, some add-in choices lead to dialog boxes that ask the user to enter information and make selections; others generate new objects or results directly.

For example, the Data Analysis ToolPak add-in inserts the choice Data Analysis into the preexisting Tools menu; the PHStat add-in inserts the PHStat pull-down menu into the menu bar. (See Figure A1.2.2) Selecting Tools | Data Analysis causes the ToolPak add-in to display the Data Analysis dialog box, in which the statistical analysis of interest can be selected. Selecting PHStat from the menu bar pulls down a menu of choices (as in Figure A1.2.2) from which the operation of interest can be selected. These selections then lead to one or more dialog boxes or to the generation of some workbook object.

Many separately obtained add-ins, including PHStat, can also be loaded temporarily by double-clicking the **PHStat for Excel 97/2000** or the **PHStat for Excel 95** icons on the Desktop (per the Installation instructions for PHStat in Appendix D) or by using the procedure to open an Excel workbook file. (Selecting PHStat from the Prentice Hall Add-Ins Start menu loads the add-in under this latter method). When loaded temporarily, add-ins can trigger the Microsoft macro virus dialog box that warns of the possibility of viruses. Should this dialog box appear, click the **Enable Macros** button to allow a virus-free add-in (such as PHStat) to be loaded. Care should be taken never to load an add-in temporarily that is already permanently installed, because unpredictable results can occur.

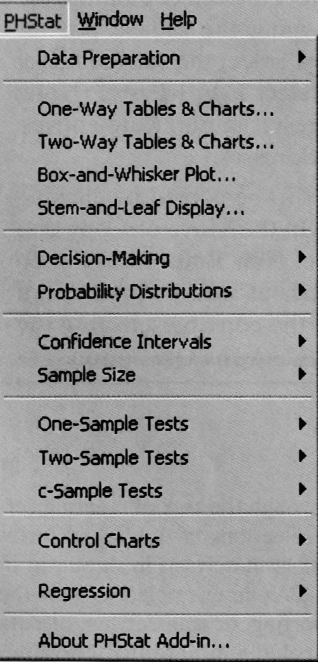


Figure A1.2.2 PHStat menu

An important reminder: to make use of Microsoft Excel and the PHStat add-in in this text, you will need to install the Analysis ToolPak and Analysis ToolPak-VBA add-ins that are supplied with Microsoft Excel and set up and install (or temporarily load) the Prentice Hall PHStat add-in included on the CD-ROM that accompanies the text. You must install the ToolPak add-ins to use PHStat; PHStat will refuse to run if it detects that the ToolPak add-ins were not previously installed.

The World Wide Web address <http://www.prenhall.com/phstat> contains answers to frequently asked questions about PHStat. It also enables users to update to the latest version of PHStat.

Summary

In this appendix, we have provided an introduction to the basic features of the Microsoft Excel application. In the remainder of the Excel appendices in this text, we will be learning about many additional aspects of Microsoft Excel in the context of specific statistical analyses.

APPENDIX 1.3 Introduction to MINITAB

What Are Statistical Software Application Programs?

In Section 1.7, when we discussed how software could help the engineer and scientist use data to make decisions, we stated that a statistical software package named MINITAB would be illustrated in this book. Statistical software application programs contain a collection of statistical methods that help provide solutions to engineering or scientific problems. These programs allow users who are relatively unskilled in statistics to access a wide variety of statistical methods for their data sets of interest. In this book, we will be illustrating and explaining the use of the MINITAB statistical software in appendices at the ends of the chapters. The latest version of MINITAB operates in a windowing environment. Readers who are not familiar with such an operating environment should read Appendix 1.1, which explains the basic features of this environment.

Entering Data using MINITAB

There are two fundamental methods for obtaining data for use with MINITAB: entering data at a keyboard, and importing data from a file. To begin, open MINITAB to obtain a window that should look similar to Figure A1.3.1.

This window contains several features that need to be discussed. At the top of the window is a Menu bar from which you make selections that allow you to obtain statistics, display graphs, store and save data, and perform many other operations. Below the Menu bar is a Session window, which displays text output (such as tables of statistics). Below the

TABLE A1.3.1	
PULSE RATE (IN BEATS PER MINUTE) OF FIVE PATIENTS UNDERGOING AN ANNUAL PHYSICAL EXAMINATION	
Name	Pulse Rate (in beats per minute)
Allen	62
Barry	70
Diane	72
Kim	58
Susan	52

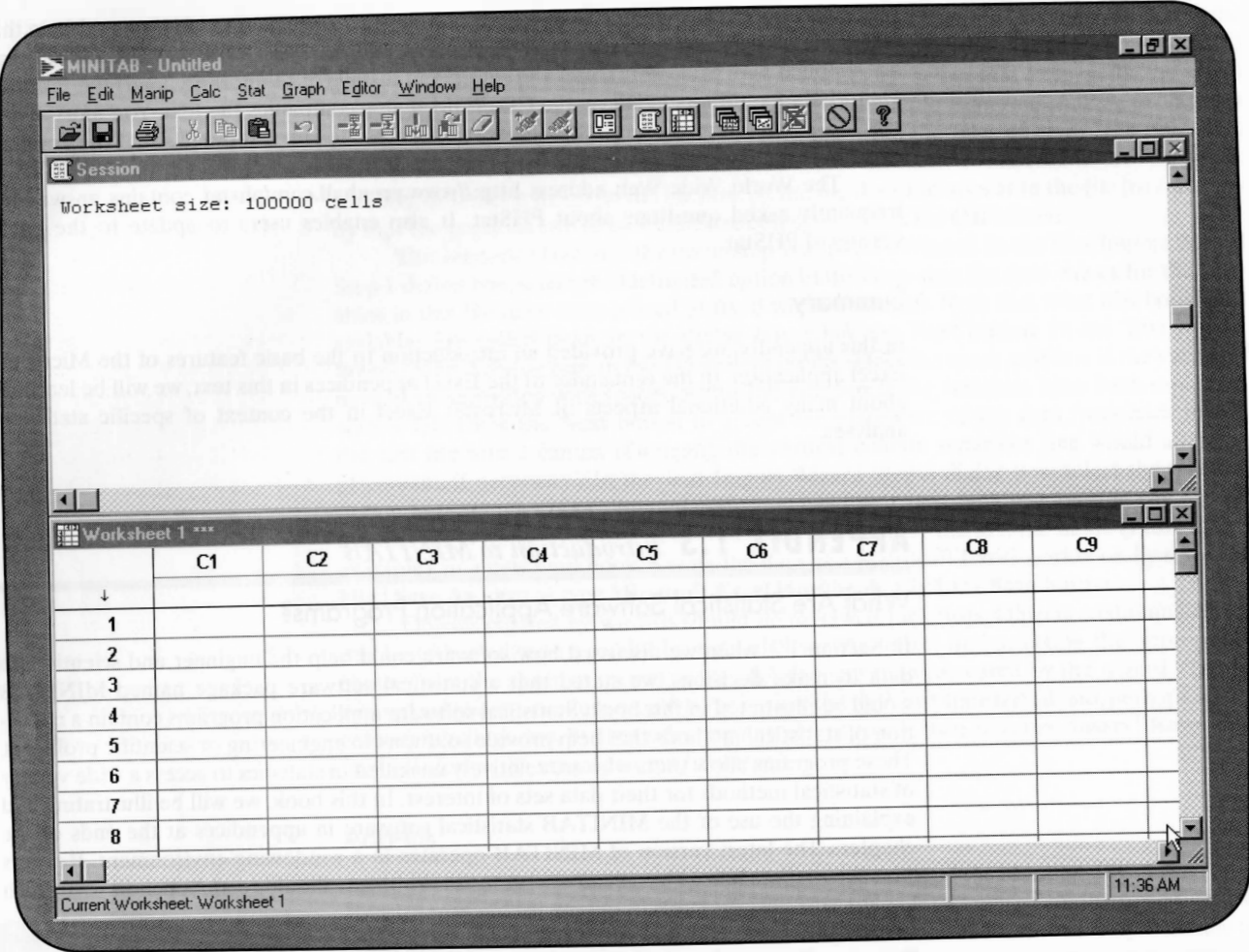


Figure A1.3.1 The MINITAB Worksheet and Session Window

Session window is the Data window, a rectangular array of rows and columns where you enter, edit, and view data. Note that the first row in the Data window lists the columns, which are generically labeled C1, C2, and so on. The following row contains an arrow in the first column and then a set of blank columns. This row serves as a place holder in which labels for variable names are entered.

Data may be entered via the keyboard in the cells in the Data window. If the arrow points down, pressing Enter moves the active cell (into which numbers are entered) down; if the arrow points right, pressing Enter moves the active cell to the right. Clicking the arrow will change the direction of the data entry.

As an illustrative example, suppose that data are available concerning the pulse rate (in beats per minute) of five patients undergoing an annual physical examination. The results are illustrated in Table A1.3.1.

Begin in the row above the horizontal line containing the numbered rows. This row is used to provide labels for each variable. In the first column (labeled C1), enter the label of the first variable (Name) and press the **Enter** key. This moves the cursor to the first row of this column. Enter Allen in row 1, Barry in row 2, Diane in row 3, Kim in row 4, and Susan in row 5. Move the cursor to the heading area at the top of column 2 (under C2), and enter Pulse as the label for this column. After pressing enter to move to row 1 in column C2, we

enter 62, the pulse rate for Allen. Continuing, enter the pulse rates for the other four patients in rows 2 to 5, respectively.

When performing statistical analyses throughout this book, we often encounter data sets with a large number of observations. When such sets have been previously entered and stored in a data file, it makes sense to try to import the contents of the file into a Data sheet, to avoid having to reenter each observation one at a time. Each data set in this text has been stored on the CD-ROM in several different formats, including the MINITAB format (.MTW). To import the contents of a MINITAB worksheet, open the MINITAB file of interest located in the appropriate directory, by selecting **File | Open Worksheet** on the menu bar and then selecting the appropriate file from the directory.

Although MINITAB can import data that have been stored in any one of various special file types (such as Microsoft Excel), it is possible that in other situations you will find data stored as text files, files that contain unlabeled and unformatted values that are separated by such delimiters as spaces, commas, or tab characters. To open data from a file, we need to use the command **File | Open Worksheet**. This will provide us with the Open Worksheet dialog box. Many different files, including MINITAB files (.MTW), Microsoft Excel (.XLS), data (.DAT), and text (.TXT), can be opened from this dialog box. Making sure that the proper file type appears in the List of Files of Type Box, select the file that you wish to open. To see how the file will look on the worksheet, click the Preview option.

If the first row of the file begins with data and does not contain the variable names (as is the case with the files ending in .TXT that are contained on the CD-ROM that accompanies this book), click the **Options** box. In the Variable Names group, select the **None** option button. In the Field Definitions group, select the **Free Format** options button. Click the **OK** button. When you are ready to open the selected file, click **Open** to open the selected file, so that it will appear in the Data window. Enter labels for each variable as appropriate, and save the file as a MINITAB worksheet by selecting **File | Save Worksheet As**.