INTRODUCTORY BUSINESS STATISTICS WITH COMPUTER APPLICATIONS

Second Edition

Adams

SECOND EDITION

Introductory Business Statistics with Computer Applications

RONALD E. SHIFFLER
ARTHUR J. ADAMS
University of Louisville

Duxbury Press
An Imprint of Wadsworth Publishing Company
A division of Wadsworth, Inc.

Editor *Curt Hinrichs*
Assistant Editor *Jennifer Burger*
Editorial Assistant *Michelle O'Donnell*
Designer *Cloyce Wall*
Print Buyer *Randy Hurst*
Copy Editor *Carol Reitz*
Technical Illustrator *Carl Brown*
Cover Designer *Jeanne Calabrese*
Cover Photo *John Still*
Signing Representative *John Douglass*
Compositor *Interactive Composition Corporation*
Printer *Banta Corporation*

This book is printed on acid-free recycled paper

ITP

International Thomson Publishing
The trademark ITP is used under license

Printed in the United States of America

2 3 4 5 6 7 8 9 10—99 98 97 96

Library of Congress Cataloging-in-Publication Data

Shiffler, Ronald E.
Introductory business statistics with computer applications / by Ronald E. Shiffler, Arthur J. Adams. — 2nd ed.
p. cm. — (The Duxbury series in statistics and decision sciences)
Prev. ed. cataloged under the title: Introductory business statistics with microcomputer applications.
ISBN 0-534-20946-7
1. Commercial statistics. 2. Commercial statistics—Computer-assisted instruction. 3. Microcomputers. I. Adams, Arthur J. II. Shiffler, Ronald E. Introductory business statistics with microcomputer applications. III. Title. IV Series.
HF1017.S5 1994
519.5'02433—dc20 93-34550

Ronald E. Shiffler is Chairman and Professor, Department of Management, University of Louisville. He has won teaching awards at the University of Florida, Georgia State University, and the University of Louisville. In 1979 and again in 1984, Dr. Shiffler won the Presentation Award in the Section on Statistical Education at the American Statistical Association's annual meeting. His research has appeared in the *American Statistician, Teaching Statistics, International Journal of Mathematical Education in Science and Technology, International Journal of Operations and Production Management,* and *Journal of Marketing Research.* He is a member of the Decision Sciences Institute, the Academy of Management, and the American Statistical Association (ASA). He formed and currently serves as President of the Kentucky chapter of the ASA. Dr. Shiffler holds a Ph.D. in Statistics from the University of Florida, a masters degree in Mathematics from Bucknell University, and an undergraduate degree from the University of North Carolina at Greensboro.

Arthur J. Adams is Professor of Business Statistics at the University of Louisville. His undergraduate degree is in Chemistry (Knox College, in Illinois), and he spent two years as a production and quality engineer in the Saran plants of Dow Chemical Company, Midland, Michigan. Dr. Adams taught at Illinois State University prior to obtaining a Ph.D. in Applied Statistics at the University of Iowa. He has published statistics-related articles in journals such as *Journal of Marketing Research, Journal of Advertising, Journal of Advertising Research, Journal of the Academy of Marketing Science, Journal of Vocational Behavior,* and *Educational & Psychological Measurement.* Dr. Adams is a member of Decision Sciences Institute, the American Marketing Association, and the American Statistical Association, and is a regular participant at the DSI National Conference. He has a daughter in high school and a son in middle school.

To the Instructor

Feedback from the first edition has been gratifying. We have received many comments from instructors saying the text is the most readable and student-oriented text available. Accordingly, we have kept our intuitive writing style as well as the emphasis on active learning via numerous examples and illustrations. We require the student to actively participate in the learning process, and show that statistics is something you do—not just something you read about. We have also incorporated into the second edition many suggestions from instructors who used the first edition, resulting in a current treatment that we believe you will wish to consider.

New in the Second Edition

We have improved the text in the following ways:

Additions

- A new software program, StataQuest, may be packaged with new copies of the text for a nominal price. Based on the commercial software STATA, StataQuest is menu-driven and has high-resolution graphics. This will be an attractive option if you wish to integrate the personal computer into your course. StataQuest is available in IBM, Macintosh, and Unix formats.
- A "full-service" chapter on quality control, Chapter 18 includes discussion and development of p-charts and c-charts in addition to the more common X-bar and R-charts. Process capability is also emphasized.
- Presentation of stem and leaf displays, box plots, sample covariance, probability counting rules, Bayes' Theorem, and index numbers.
- Instructions and examples for SAS and MINITAB, both the command and menu-driven versions, to generate various statistical analyses. In addition, most chapters have several exercises that ask the student to interpret SAS and MINITAB printouts.
- A new feature called "Learning with Live Data" at the end of most chapters. These are activities where the students in the class are individually responsible for generating the data. The instructor or the students can then collate the results and distribute them to the class with a specific assignment either as an individual or team activity. These exercises are intended to highlight problem definition as well as to illustrate difficulties in collecting and organizing real data, while at the same time reinforcing chapter topics. Extensive notes to the instructor on using "Learning with Live Data" sets appear in the *Instructor's Resource Manual.*

- Two data set appendixes. Appendix P contains eight different time series (two monthly, two quarterly, and four annual) related to the U.S. economy. Appendix N contains information on eight variables for 112 residential real estate properties; this data set may be drawn upon while teaching in the regression chapters and elsewhere.
- We have identified the exercises in each chapter with various icons to signify potential applications in business and/or to denote the problem type. The icons used are shown below:

Other changes

- We have streamlined the inference chapters, combining chapters on large sample inference, small sample inference, and inferences for π into one chapter.
- We have added, changed, or updated over 250 exercises. With about 1500 exercises, we offer more problems for the instructor to choose from than most, if not all, other books.

Optional Customization

For schools wishing to use a book that includes some, but not all, of the 18 chapters in this text, Duxbury Press will print and bind a shorter version. The cost to the student of such a book will often be less than the full book—the exact cost will depend on the amount of material included and the number of students taking the course. Please contact your Wadsworth representative for further details. On the next page we give a few of many possible chapter outlines an instructor might consider using in a one-term course.

	DESCRIPTION	CHAPTERS
Example 1	A traditional course with a brief exposure to simple linear regression	1–10
OPTION: Time permitting, add Chapter 11 on two-sample inference.		
Example 2	A traditional course with a solid grounding in regression	1–3, 5–10, 13
OPTION: Time permitting, add Chapter 11 on two-sample inference.		
Example 3	A nontraditional course with reduced emphasis on probability	1–3, 6–11
OPTION: Time permitting, add Chapter 4 or Chapter 13 on regression, or add Chapter 18 on quality control.		

Chapters 12–17 deal with topics usually covered in a second course on business statistics. With the exception of Chapter 13, we have not suggested that any of these chapters be included in a one-semester course. However, you can easily include one or more of them in a one-semester course if you desire, without any loss of continuity.

Other Special Features

Classroom Practice Sprinkled throughout the chapters are problems labeled "Classroom Practice." These multi-part problems can be used in several ways. One possibility is this: After the instructor has introduced a new topic (such as how to use the binomial formula and tables), the students may be assigned to work a Classroom Practice that bears directly on the new topic. The students can work in-class on the problem either individually or in small groups. Working as a coach, the instructor is then free to walk about the class and assist any student who needs help. After perhaps seven to ten minutes, the instructor can go over the solution with the whole class. We think the students benefit from getting immediate experience working a problem right after its presentation by the instructor. In addition, the instructor can assess the class's comprehension of the material.

Learning with Live Data As described above, this innovative problem format involves the students in generating the sample observations, thereby increasing their involvement. These activities (which may be assigned to groups or teams if desired) are less structured than typical textbook exercises, and often do not yield a single "correct" answer, thereby increasing the opportunities for class discussions.

"To Be Continued . . ." Sections At the conclusion of most chapters is a short section titled "To Be Continued. . . ." The specific purpose of these sections is twofold:

1 To alert the student to the fact that material just studied may be encountered again in future college courses (although perhaps with a slightly different

name or symbol). We provide several connections to finance, management science, marketing, and accounting.

2 To alert the student to applications of chapter material in the media as well as in business practice.

More generally, the idea of "To Be Continued..." is to refute the notion that business statistics doesn't relate to other disciplines or that it is not useful. We think it is important to show where applications exist. To this end, we often cite press sources such as *USA Today*, *Business Week*, *Los Angeles Times*, and so forth. In addition, several statistics-related questions are reproduced verbatim from past CPA exams. Our intent is to provide ample opportunity for students to see applications of business statistics in the real world.

Section on Opinion Polling As an extension of the idea of showing applications of business statistics, we offer a section (in Chapter 10) on opinion polling. We relate polling to making inferences about the binomial parameter and expand on the need for a random sample taken from a clearly defined target population. This section has worked well in the first edition, and there are always current polls in the news that can be used as fresh examples of well or poorly conducted or reported cases for class discussion.

Wide Problem Selection This book has over 1500 problems. Each chapter has a variety of exercises that range from drill work to more challenging problems. We provide exercises from a variety of scenarios. Most chapters have several problems that are taken from current literature, including various research journals, such as the *Journal of Marketing Research*, *American Statistician*, and *Journal of Advertising Research*. Many other problems make use of "real" data sets. We use optional exercises to denote problems involving topics or finer points that we did not have space to discuss as a regular part of the chapter. Some examples of concepts treated in this fashion include the geometric mean, the hypergeometric distribution, the uniform distribution, a one-sided confidence interval, and various alternative computing formulas not usually seen.

Computer Support Our book can be used in conjunction with either mainframe or PC support. As mentioned above, the text offers examples of commands and output for MINITAB (both the line command and the menu-driven versions) and SAS for all major applications. For instructors interested in a more extensive PC package with extended graphics capabilities, the student edition of EXECUSTAT 3.0, JMP-IN©, and STATA are available for sale to students from the publisher. For a basic program, the StataQuest program mentioned above can be ordered with new copies of the text.

Content and Flexibility

The first ten chapters cover what many instructors consider a proper amount of material for a first course in statistics. Instructors teaching a second term are likely to vary their coverage beyond this point to suit their own preferences and objectives.

Chapters 1–3 concentrate on problem structure, terms, organizing data, and descriptive statistics. We place some emphasis on the Empirical Rule in Chapter 3, using the ± 2 standard deviations statement as a preliminary device for setting up the notion of rejection regions later on in the course.

Chapter 4 discusses the ideas of correlation, covariance, and the least-squares line in a descriptive context. This brief treatment is intended to ensure that those students who will take only one statistics course in their career will have been exposed to these important concepts.

Chapter 5 introduces probability basics, such as types of probability, counting rules, and Bayes' Theorem. In Chapters 6 and 7, we cover specific probability distributions: the binomial, Poisson, normal, and exponential. While we would expect all instructors to teach the binomial and normal distributions, personal taste and/or time constraints may influence the inclusion of the others.

Chapter 8 deals specifically with the sampling distribution of the mean and the proportion, although it lays out the logic for developing other sampling distributions. Section 8.3 treats the case of sampling from finite populations and may be viewed as optional. After this chapter the student will be ready to consider statistical inference.

Chapter 9 is a short transitional chapter that sets forth the logic, framework, and vocabulary of inferential statistics. None of the problems posed in Chapter 9 require a calculator for solution.

Chapter 10 develops confidence intervals for μ and hypothesis tests about the value of μ for large samples; then the t-distribution is introduced for samples less than size 30. The chapter concludes with a section on inferences for the binomial parameter π and a section on polling.

Chapter 11 covers the case of two-population sampling, including the concept of matched pairs. Chapter 12 introduces ANOVA and the F-distribution. In addition to one-way analysis of variance, we discuss follow-up tests, the randomized block design, and two-factor ANOVA with interaction. Chapter 13 is the first of an extensive three-chapter regression sequence. Chapter 13 offers the simple regression and correlation models, while Chapter 14 generalizes to the multiple regression case. Chapter 15 is concerned with model building for multiple regression.

Chapter 16 investigates forecasting for business and economic time series. A variety of forecasting methods are presented along with the concept of combining forecasts as a strategy for forecast improvement. Chapter 17 presents chi-square tests and several of the more useful nonparametric statistical methods.

Chapter 18 is a flexible chapter on quality control. It may be covered at almost any point desired after Chapter 8.

Supplemental Materials

- The *Instructor's Resource Manual* contains the following: complete solutions (not just answers) to all text problems, a test bank including objective and computational questions, a set of transparency masters for classroom use (primarily selected from figures in the text), and instructor's notes for using the "Learning with Live Data" exercises. Computer-related exercises for the regression chapters include complete SAS and MINITAB printouts. The *Instructor's Resource Manual* was prepared by the authors in its entirety. Computerized testing is available from the publisher in both IBM and Macintosh versions.
- A data disk is included with each new copy of the book and contains almost all the exercises that use raw data sets, the Classroom Practice problems, as well as the more extensive data sets that make up Appendixes N and P.

Overall, we think that the intuitive orientation of the text and the emphasis on real applications, when combined with the supporting materials, will enhance your classroom teaching and your students' appreciation of business statistics.

Acknowledgments

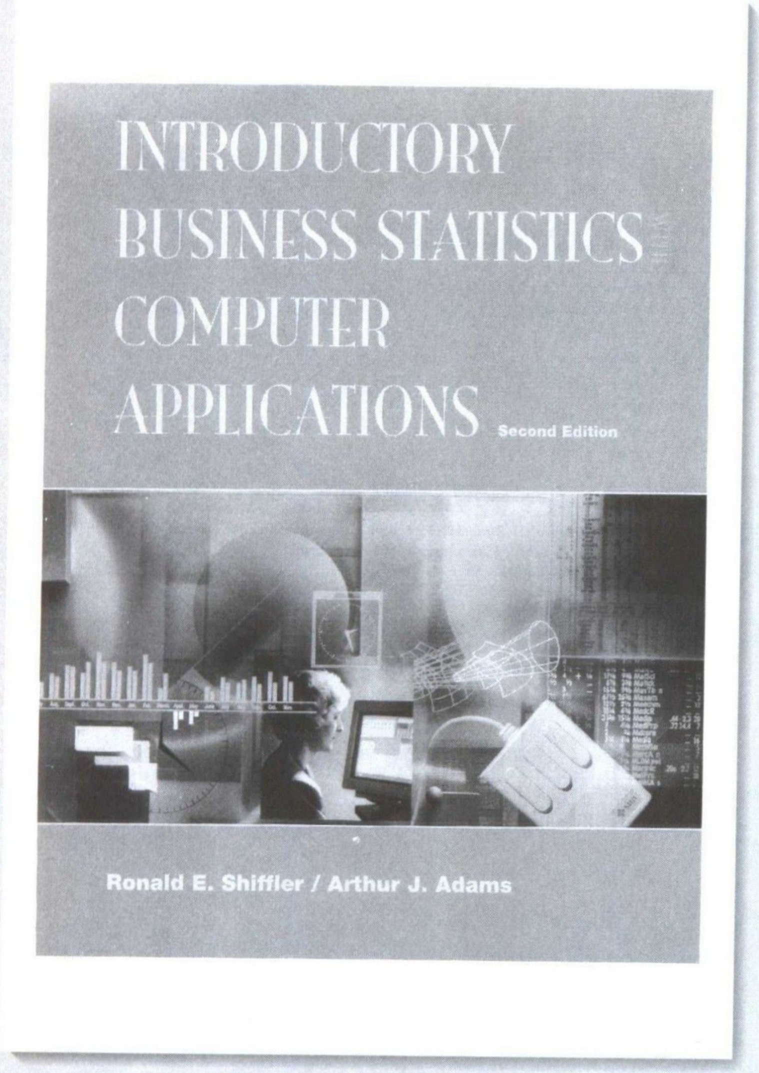
Creating a book is a team effort and we would like to acknowledge our team members here. At the head of our team is our coach and editor, Curt Hinrichs, who provided support, guidance, and experience at all times. Assistant editor Jennifer Burger and production coordinator Phyllis Niklas had the responsibility for turning our manuscript into a book. We feel fortunate to have worked with these professionals.

We would like to thank the following reviewers (listed in alphabetical order) for their constructive comments, which helped improve this second edition: Priscilla Chaffe-Stengel, California State University, Fresno; Kimon Constas, Fayetteville State University; Rick L. Edgeman, Colorado State University; Herb Hooper, Chattanooga State Technical Community College; Robert W. Hull, Jr., Western Illinois University; Max Jerrell, Northern Arizona University; Dennis Kern, University of Texas, San Antonio; Ronald Klimberg, Boston University; Jim Lowber, General Electric; Jeff Mock, Diablo Valley College; Joseph J. Moder, University of Miami; Bob Myers, University of Louisville; Don R. Robinson, Illinois State University; and Jack Wilson, North Carolina State University. We also thank Missy Mountz and Jennifer Young for their help in running MINITAB and SAS, checking answers to some exercises, and preparing permissions materials.

In addition, we would like to acknowledge the cooperation of those individuals and organizations who allowed us to reprint selected material.

Ronald E. Shiffler
Arthur J. Adams

Your Visual Guide to This Book



In this book you will discover that learning business statistics is based on five separate and equally important elements:

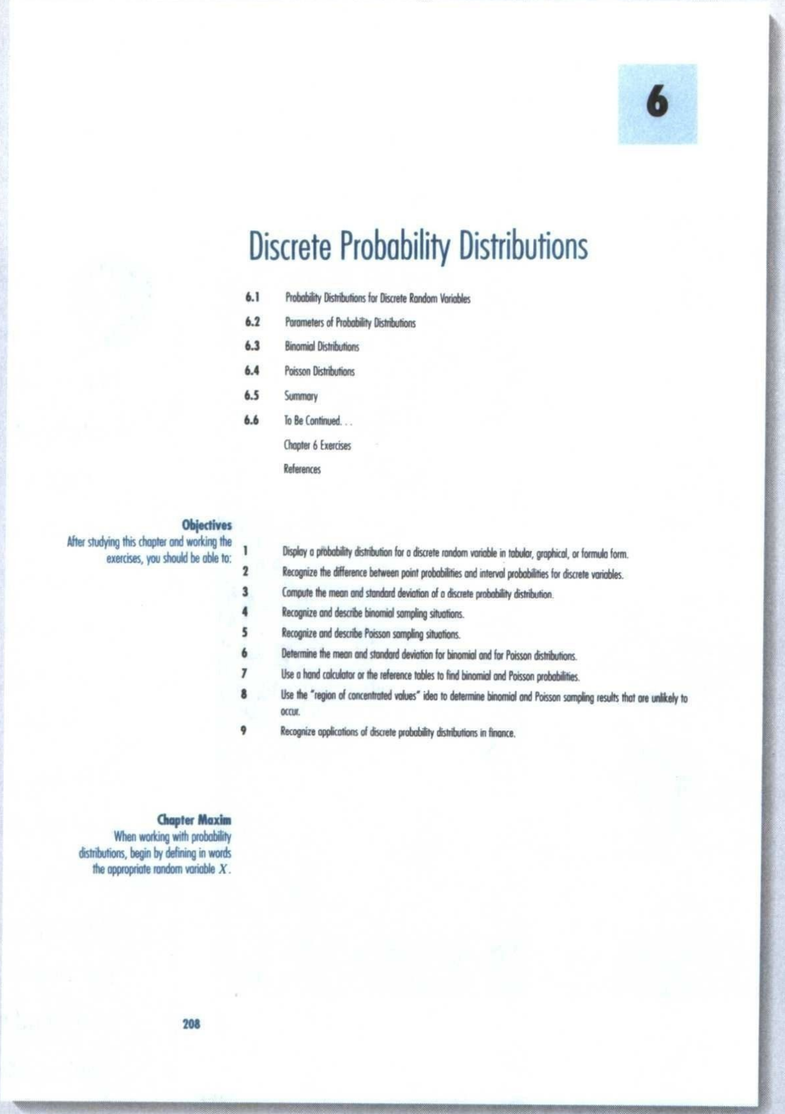
- *planning*
- *execution*
- *reporting*
- *evaluation*
- *action*

These components provide an effective plan for using this book, too. By applying the five-step process as you read, generate and interpret computer data, work problems, and review, you will learn and retain the principles that govern use of statistics in business today.

This guide acquaints you with features in the book specifically designed to extend your grasp of statistical concepts and procedures, thereby helping you succeed in your course today and in your business career tomorrow.

P L A N N I N G

A list of **objectives** focuses your attention on the purpose of the chapter. Keep these goals in mind as you read assigned sections and do the exercises. Then, when you review, check back to determine how well you have accomplished all of the objectives. These objectives may be especially useful as you study for a test because many test questions are designed to measure comprehension of each learning objective.



6

Discrete Probability Distributions

Objectives

After studying this chapter and working the exercises, you should be able to:

1. Display a probability distribution for a discrete random variable in tabular, graphical, or formula form.
2. Recognize the difference between point probabilities and interval probabilities for discrete variables.
3. Compute the mean and standard deviation of a discrete probability distribution.
4. Recognize and describe binomial sampling situations.
5. Recognize and describe Poisson sampling situations.
6. Determine the mean and standard deviation for binomial and for Poisson distributions.
7. Use a hand calculator or the reference tables to find binomial and Poisson probabilities.
8. Use the "region of concentrated values" idea to determine binomial and Poisson sampling results that are unlikely to occur.
9. Recognize applications of discrete probability distributions in finance.

Chapter Maxim

When working with probability distributions, begin by defining in words the appropriate random variable X.

208

A chapter **maxim** presents an essential idea to motivate and direct your study of material presented in the chapter.

A number of devices help you organize your study and review: **Comments,** set off by ruled lines, clarify and expand concepts covered in the text.

Rules, principles, and formulas are boxed off for emphasis. These "call-outs" direct your attention to important subject matter and also provide a way to locate it quickly as you solve problems or review material.

Definitions, clearly identified, are also boxed for easy reference.

COMMENT You cannot always determine whether quantitative data are discrete or continuous by looking at the numbers. Because of space and accuracy limitations, continuous data always are rounded and therefore look like and are even treated like a set of discrete numbers. For example, the time required to close a sales deal is, in theory, continuous but reported discretely, such as 83 minutes, 83.2 minutes, and so on. Although we cannot list all the trailing decimals for continuous data, this does not nullify the continuous nature of the data.

Principles of Inclusion and Exclusion

Another basic tool for organizing data is to use the two straightforward rules given in the following box:

Principle of Inclusion

There must be a category into which each piece of data can be assigned.

Principle of Exclusion

No piece of data can be assigned to more than one category.

Simply stated, the Principles of Inclusion and Exclusion require each piece of data, whether qualitative or quantitative, to fit into one and only one category. We do not wish to exclude any pieces of data, nor do we wish to "double count" any data points.

We use the word *category* in a generic sense. For qualitative data, the term is self-explanatory. For quantitative data, a category is either an actual value of the variable or a *class* of consecutive values. To make it easier to organize data, we will adopt a special definition for the term *class*.

DEFINITION ▪ A **class** that begins at point a and ends at point b is a set of consecutive values that are greater than or equal to a, but less than b. ▪

For example, the class from 45 to 55 for the variable X means $45 \le X < 55$.

EXAMPLE **2.2** Develop several categories to account for the possible values for these variables:

a Number of times per week an individual eats the evening meal at a restaurant

b Interest rate charged on bank credit cards

Solution

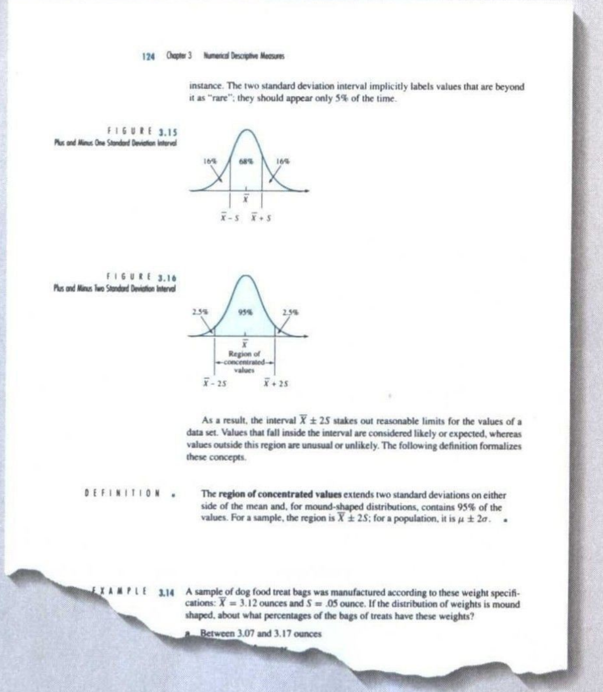
a If we assume that an individual eats only one evening meal per day, then the possible categories are the integers from 0 to 7. In this case the categories are the actual values of the variable.

b The minimum value would be 0%, but the maximum value is unknown. We might expect a ceiling figure to be around 24%. With this in mind, some potential classes are listed in the margin.

Classes (%)
0 – under 6
6 – under 12
12 – under 18
18 – under 24

E X E C U T I O N

The text is filled with opportunities to learn by doing: **Worked examples** (between 10 and 20 per chapter) present typical problems based on subjects just covered in the text. For lasting grasp of concepts and techniques, it is best to work these examples yourself, step-by-step, checking your answers—and your understanding of the rationale behind them—with the fully worked-out solutions that are supplied.


124 Chapter 3 Numerical Descriptive Measures

instance. The two standard deviation interval implicitly labels values that are beyond it as "rare"; they should appear only 5% of the time.

FIGURE 3.15
Plus and Minus One Standard Deviation Interval

FIGURE 3.16
Plus and Minus Two Standard Deviation Interval

As a result, the interval $\overline{X} \pm 2S$ stakes out reasonable limits for the values of a data set. Values that fall inside the interval are considered likely or expected, whereas values outside this region are unusual or unlikely. The following definition formalizes these concepts.

DEFINITION ▪ The **region of concentrated values** extends two standard deviations on either side of the mean and, for mound-shaped distributions, contains 95% of the values. For a sample, the region is $\overline{X} \pm 2S$; for a population, it is $\mu \pm 2\sigma$. ▪

EXAMPLE 3.14 A sample of dog food treat bags was manufactured according to these weight specifications: $\overline{X} = 3.12$ ounces and $S = .05$ ounce. If the distribution of weights is mound shaped, about what percentages of the bags of treats have these weights?

a Between 3.07 and 3.17 ounces

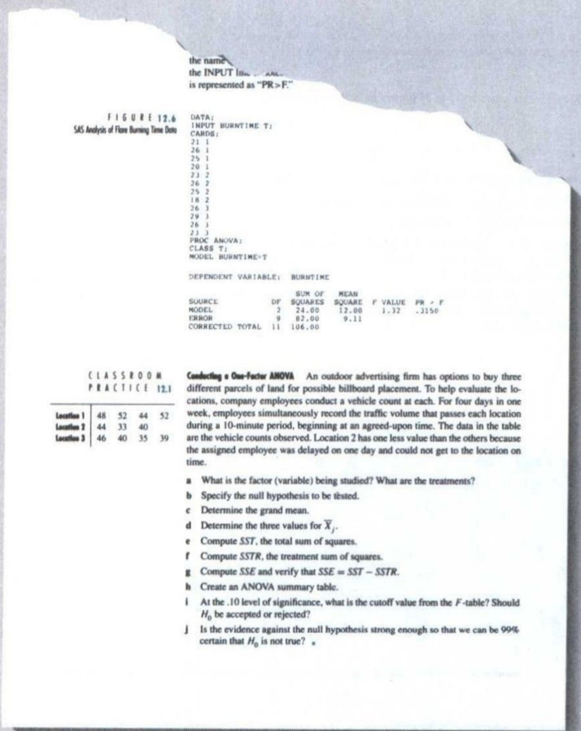

the name
the INPUT line
is represented as "PR>F."

FIGURE 12.6
SAS Analysis of Flare Burning Time Data

```
DATA;
INPUT BURNTIME T;
CARDS;
21 1
26 1
25 1
20 1
23 2
26 2
25 2
18 2
26 3
29 3
26 3
23 3
PROC ANOVA;
CLASS T;
MODEL BURNTIME=T

DEPENDENT VARIABLE:  BURNTIME

                        SUM OF    MEAN
SOURCE             DF  SQUARES   SQUARE  F VALUE  PR > F
MODEL               2   24.00    12.00     1.32   .3150
ERROR               9   82.00     9.11
CORRECTED TOTAL    11  106.00
```

CLASSROOM PRACTICE 12.1

Location 1	48	52	44	52
Location 2	44	33	40	
Location 3	46	40	35	39

Conducting a One-Factor ANOVA An outdoor advertising firm has options to buy three different parcels of land for possible billboard placement. To help evaluate the locations, company employees conduct a vehicle count at each. For four days in one week, employees simultaneously record the traffic volume that passes each location during a 10-minute period, beginning at an agreed-upon time. The data in the table are the vehicle counts observed. Location 2 has one less value than the others because the assigned employee was delayed on one day and could not get to the location on time.

a What is the factor (variable) being studied? What are the treatments?
b Specify the null hypothesis to be tested.
c Determine the grand mean.
d Determine the three values for $\overline{X}_j$.
e Compute *SST*, the total sum of squares.
f Compute *SSTR*, the treatment sum of squares.
g Compute *SSE* and verify that $SSE = SST - SSTR$.
h Create an ANOVA summary table.
i At the .10 level of significance, what is the cutoff value from the F-table? Should H_0 be accepted or rejected?
j Is the evidence against the null hypothesis strong enough so that we can be 99% certain that H_0 is not true? ▪

Classroom practice exercises, located at the end of most sections, are intended to provide immediate feedback on how well you have absorbed material just covered in your instructor's lectures.

R E P O R T I N G

There are many ways of expressing the statistical ideas and procedures used in the business environment. Words form the main part of the text, of course. (That is why it is so important to read sections as they are assigned.) But, often, subject matter is easier to comprehend when it is presented symbolically. You will find graphics carefully integrated with the text in this book:

226 Chapter 6 Discrete Probability Distributions

- Number of potholes per mile of city street

At first the Poisson may seem somewhat like the binomial, but brief consideration of the circumstances should tell you which distribution is appropriate. Table 6.2 provides a comparison of these two distributions. The Poisson and binomial are alike in that they are discrete distributions where the random variable can take on only positive integer values or 0. One major difference is the absence of the sample size concept in the Poisson distribution. The binomial is used to count occurrences (successes) *for a given sample size*, whereas the Poisson counts occurrences *within a given interval of time*, area, and so forth. Expressed another way, when using the binomial, we can count either of two opposite outcomes: the *successes* or the *failures* in a given number of entities. But with the Poisson, we can count only one outcome: the number of *occurrences* of some event of interest. Referring to some of the Poisson examples given above, we easily can count the number of vehicles arriving at a toll bridge, but it makes no sense to speak of the opposite (the number that do not arrive). Likewise, we can count lost-time accidents as they happen, but it is meaningless to speak of the number of accidents that did not happen.

TABLE 6.2
Comparing the Binomial and Poisson Distributions

Similarities	Poisson	Binomial
Type of distribution	Discrete	Discrete
Random variable X	Number of occurrences	Number of successes
Possible values of X	$X = 0, 1, 2, 3, \ldots$	$X = 0, 1, 2, 3, \ldots$, up to n
Differences		
Number of entities (sample size) concept	Absent	Present
Interval of time (or other measurement) concept	Present	Absent
Countable outcomes	One	Two

Like the binomial distribution, the Poisson requires some specific conditions to be met:

Poisson Sampling Conditions

1 We are counting the number of times some event happens within some defined continuous interval.
2 The number of occurrences in any one interval is independent of the number in any other nonoverlapping interval.
3 The probability of an occurrence is the same in all intervals of length.

Notice that th

Tables organize material visually, making it possible to track a number of elements in different dimensions—especially those that would be difficult to follow verbally.

40 Chapter 2 Organizing and Graphing Data

recheck the origin and validity of an unusual response. If we find nothing in error, then a workable solution is to isolate (temporarily) the data point, organize the remaining $n - 1$ observations, and then include the outlier in a separate category.

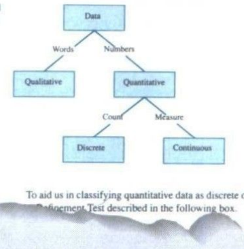
These and other activities, such as coding the data or creating a data file for later use with a computer program, are considered necessary parts of managing data. We will use the term *processed* to describe the raw data *after* the various data management activities have been performed. A more complete treatment of data management techniques usually occurs in a marketing research course or in an advanced statistical methods class, so we will not delve into this area in more detail here. We shall proceed on the assumption that our data sets have been "cleaned up" as well as possible. Before considering ways to organize and visually present processed data, we need to develop some ground rules.

Two Types of Quantitative Data

In Chapter 1 we learned that sampling produces one of two forms of data: qualitative and quantitative. Although qualitative data can be thought of as one broad category, we will find it useful to break down quantitative data into two types: discrete and continuous. *Discrete* means separate and distinct; *continuous* means unbroken or extending without interruption. A quantitative variable that has "separate" values at specific points along the number line, with gaps between them, is called a **discrete variable**. A quantitative variable that has a "connected" string of possible values at all points along the number line, with no gaps between them, is called a **continuous variable**.

Deciding whether quantitative data are values of a continuous variable or a discrete variable is sometimes difficult. In Figure 2.1 we have indicated a characteristic of each type of data along the lines that connect the boxes. A discrete variable usually produces data by counting. For example, counting the number of employees in a company yields discrete data. Continuous data usually result from a measurement or computation. For instance, dividing the number of unemployed people by the size of the work force, we generate continuous data for the variable *unemployment rate*. Continuous variables are also characterized by physical measurements such as size, time, area, and so forth.

FIGURE 2.1
Data Classifications and Characteristics



To aid us in classifying quantitative data as discrete or continuous, we may resort ... Test described in the following box.

Charts and graphs provide another way to make processes and relationships among data evident. They also demonstrate techniques you can duplicate to communicate your own statistical findings.

E V A L U A T I O N

Computers are a fact of life in business and business statistics today. This book prepares you to participate in that environment with consistent reference to computer data and output, using the most popular "real-world" statistical programs.

(Containing a data file for most exercises and classroom practices, the data disk included with this book provides the information to create your own computer output, applied to the sort of problems you will encounter in actual business practice.)

Reproductions of **computer screens** are accompanied with step-by-step instructions you can use to generate and interpret data.

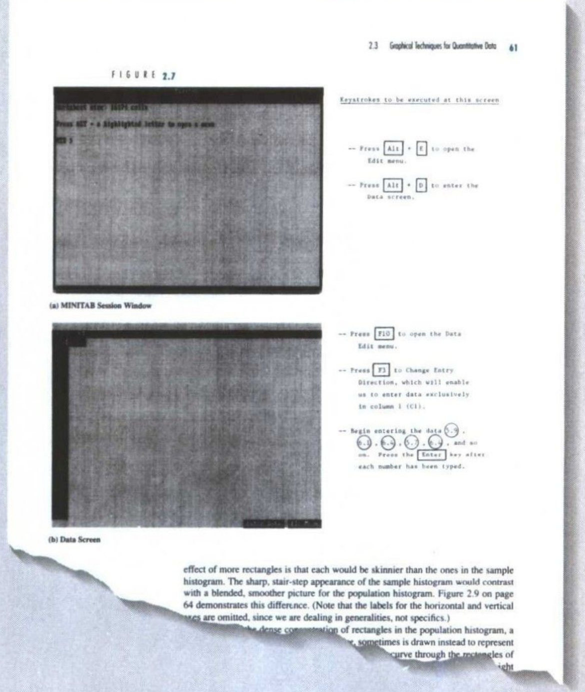

2.3 Graphical Techniques for Quantitative Data 61

FIGURE 2.7

Keystrokes to be executed at this screen

-- Press Alt + E to open the Edit menu.

-- Press Alt + D to enter the Data screen.

(a) MINITAB Session Window

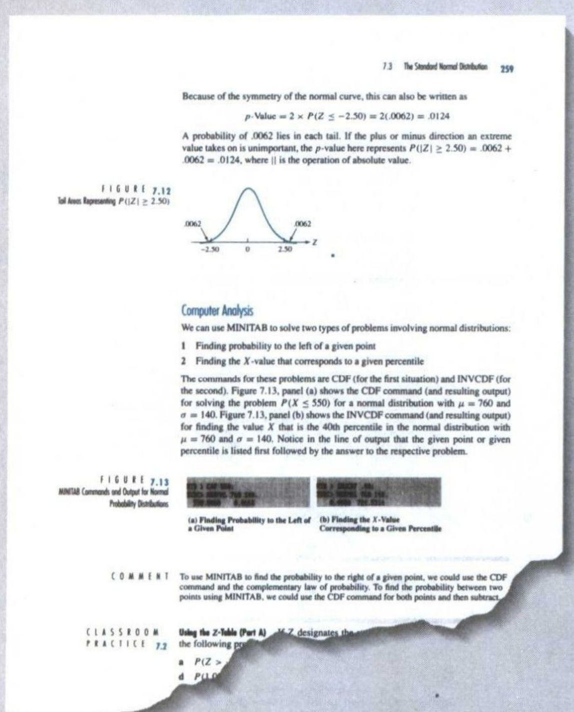
-- Press F10 to open the Data Edit menu.

-- Press F3 to Change Entry Direction, which will enable us to enter data exclusively in column 1 (C1).

-- Begin entering the data 5.9, 6.1, 6.4, 5.7, 6.4, and so on. Press the Enter key after each number has been typed.

(b) Data Screen

effect of more rectangles is that each would be skinnier than the ones in the sample histogram. The sharp, stair-step appearance of the sample histogram would contrast with a blended, smoother picture for the population histogram. Figure 2.9 on page 64 demonstrates this difference. (Note that the labels for the horizontal and vertical axes are omitted, since we are dealing in generalities, not specifics.)


7.3 The Standard Normal Distribution 259

Because of the symmetry of the normal curve, this can also be written as

$$p\text{-Value} = 2 \times P(Z \le -2.50) = 2(.0062) = .0124$$

A probability of .0062 lies in each tail. If the plus or minus direction an extreme value takes on is unimportant, the p-value here represents $P(|Z| \ge 2.50) = .0062 + .0062 = .0124$, where || is the operation of absolute value.

FIGURE 7.12
Tail Areas Representing $P(|Z| \ge 2.50)$

Computer Analysis

We can use MINITAB to solve two types of problems involving normal distributions:

1 Finding probability to the left of a given point
2 Finding the X-value that corresponds to a given percentile

The commands for these problems are CDF (for the first situation) and INVCDF (for the second). Figure 7.13, panel (a) shows the CDF command (and resulting output) for solving the problem $P(X \le 550)$ for a normal distribution with $\mu = 760$ and $\sigma = 140$. Figure 7.13, panel (b) shows the INVCDF command (and resulting output) for finding the value X that is the 40th percentile in the normal distribution with $\mu = 760$ and $\sigma = 140$. Notice in the line of output that the given point or given percentile is listed first followed by the answer to the respective problem.

FIGURE 7.13
MINITAB Commands and Output for Normal Probability Distributions

(a) Finding Probability to the Left of a Given Point
(b) Finding the X-Value Corresponding to a Given Percentile

COMMENT To use MINITAB to find the probability to the right of a given point, we could use the CDF command and the complementary law of probability. To find the probability between two points using MINITAB, we could use the CDF command for both points and then subtract.

CLASSROOM PRACTICE 7.2 **Using the Z-Table (Part A)**

Computer analysis sections demonstrate how statistical packages—such as MINITAB, SAS, and Executstat—can be used in typical business-oriented situations.

FIGURE 4.28

```
                    EXECUSTAT STUDENT EDITION

Search                   Browse line: 1             Esc=Exit   F1=Help

Linear model: EstimMPG = 39.0685 - 0.0654737*EngineSize

                        Table of Estimates
                               Standard          t              P
                 Estimate       Error           Value          Value
Intercept        39.0685        1.33205         29.33          0.0000
Slope            -0.0654737     0.00592082      -11.06         0.0000
R-squared = 93.86%
Correlation coeff. = -0.969
Standard error of estimation = 1.152
Durbin-Watson statistic = 1.20897
Mean absolute error = 0.924522
```

Note: Exercises 4.56 and 4.57 are based on the 112 observations in the real estate data set that appears in Appendix N at the back of the book.

4.56 a Find the equation of a straight line that best relates X = hundreds of square feet to Y = selling price.
b Use your equation to predict the selling price of a new home on the market that has 2600 square feet.
c What is the sample correlation coefficient?

4.57 a Obtain a scatter plot that relates lot size to selling price.
b Determine the sample correlation coefficient.
c Determine the sample correlation coefficient for selling price and number of bedrooms. Which of lot size or number of bedrooms is more strongly related to selling price?

Learning with Live Data

4.58 **Snow Skiing** Advertisers sometimes take a few liberties with the facts. Consider the case of ski resort operators in New England. In a comparison of the actual number of miles of ski trails versus the number of miles advertised by the resort, all six ski resorts included in a study that was reported in the *Wall Street Journal* overestimated the truth. Is there a trend in the hyped data? By analyzing the data below, answer this question and others that your instructor may pose. Define the following variables:

X = actual miles of trails as determined by survey
Y = miles of trails claimed by the resort
$D = Y - X$ (the overstatement in miles)
$P = (D/X) \times 100\%$ (the percentage overstatement in miles)

The data are as follows:

"Learning with Live Data" problems are designed for collaborative collection and generation of "real" data in your computer lab or in the classroom. These problems accustom you to the team efforts that are often used in actual business practice today.

9.5 To Be Continued . . . 31

9.5 To Be Continued . . .

. . . In Your College Courses

The two ways of making statistical inferences—estimation and hypothesis testing—reappear in other business disciplines. For students who are accounting majors we present a brief glimpse into auditing, where these inferential tools are used extensively.

In Chapter 14 of Arens and Loebbecke's textbook titled *Auditing: An Integrated Approach*, we read how auditors draw statistical inferences. These can be either confidence intervals (for instance, to estimate the error amount when there is both a recorded value and an audited value for the sample items) or a hypothesis test (for instance, rendering a conclusion as to whether or not a client's books are materially misstated).

Arens and Loebbecke go on to discuss the risks associated with the two types of hypothesis-testing errors:

- ARIA, the auditor's acronym for the "acceptable risk of incorrect acceptance," which is the chance of committing a Type II error (accepting a population of accounts that is actually materially misstated). There are potential legal implications here.
- ARIR, the "acceptable risk of incorrect rejection," which is the chance of committing a Type I error (concluding that a population is misstated when it is not). This could mean increased sample sizes and/or additional tests involving greater costs.

Clearly a student who is pursuing a career in auditing or public accounting will encounter statistical hypothesis testing on a regular basis. Hypothesis testing is also an integral part of business courses in marketing research, econometrics, forecasting, portfolio theory, and to a lesser extent, management science and personnel management.

. . . In Business/The Media

Estimation and hypothesis testing are found in many areas of the business world, although the applications may not appear as pure as what we have described in this chapter.

Clearly, the EPA's efforts to estimate and report the gas mileage of new cars (see Example 9.2) is a pure application of statistical inference. As another example of the government's use of statistical methods, consider the New Car Assessment Program conducted by the National Highway Traffic Safety Administration (NHTSA). Each year NHTSA buys about 30 new cars and crashes them into a solid barrier at 35 miles per hour. Two dummies wired with electronic sensors are placed in the front sea[illegible] of each car crashed. The sensors generate data about the severity of the impact [illegible] are used to estimate the so-called HIC (Head Injury Criterion) sc[illegible] G's" measurement (the deceleration load at the chest) for each [illegible] report is published in the media in [illegible]

"To Be Continued" sections link the material covered in a chapter to other, more advanced courses you may take and to practical applications that you will encounter in your business career.