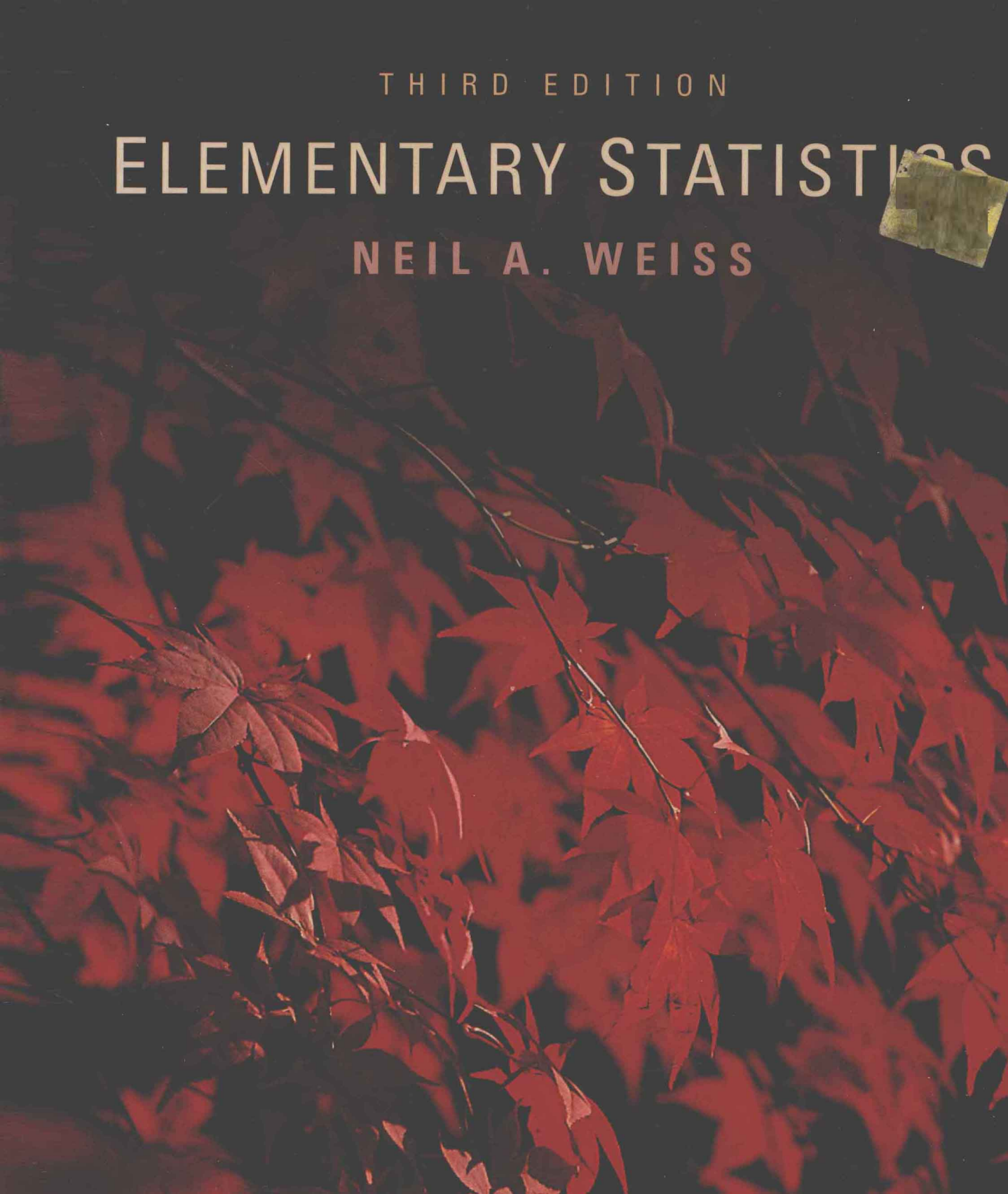THIRD EDITION

ELEMENTARY STATISTICS

NEIL A. WEISS

Elementary Statistics

• • • • •

Third Edition

Neil A. Weiss
Arizona State University

Biographies by Carol A. Weiss

Photographs:

Florence Nightingale, page 2, courtesy of The Bettmann Archive.
Adolphe Quetelet, page 42, Copyright Bibliothèque royale Albert Ier,
    Cabinet des Estampes, Bruxelles, J. Odevaére.
John Tukey, page 100, courtesy of John W. Tukey.
Andrei Kolmogorov, page 180, courtesy of The Bettmann Archive.
James Bernoulli, page 240, courtesy of The Bettmann Archive.
Carl Friedrich Gauss, page 298, courtesy of The Bettmann Archive.
Pierre-Simon Laplace, page 366, courtesy of The Bettmann Archive.
William Gosset, page 402, courtesy of The Granger Collection, New York.
Jerzy Neyman, page 444, courtesy of the Department of Statistics, University of
    California, Berkeley.
Sir Ronald Fisher, page 510, courtesy of The Bettmann Archive.
Abraham de Moivre, page 564, courtesy of The Granger Collection, New York.
Karl Pearson, page 598, courtesy of Brown Brothers.
Adrien Legendre, page 646, courtesy of The Bettmann Archive.
Sir Francis Galton, page 696, courtesy of Stock Montage.

To my father
and the
memory of my mother

Preface

. . . . .

Statistics has become an indispensable tool in business, government, and virtually every academic discipline. Some familiarity with statistics is essential for all of us if we are to comprehend the world around us.

The purpose of this book is to provide a clear understanding of basic statistical concepts and techniques and to present well-organized procedures for applying them. Introductory high-school algebra is a sufficient prerequisite.

The text is designed so that it can be used in either a one-quarter or one-semester course. The amount of time devoted to the book can be varied by both choice of topics and depth of coverage.

Technological advances and ever increasing calls for new approaches to presenting statistics have made this an exciting time to learn, practice, and teach statistics. In writing the third edition of *Elementary Statistics,* we have incorporated many of the techniques and attitudes that reflect recent developments in statistics.

Features

The text contains the following features that will provide valuable assistance for the reader in learning elementary statistics.

Emphasis on application.   We have concentrated on the application of statistical techniques to the analysis of data. Although statistical theory has been kept to a minimum, we have provided a thorough explanation of the rationale for using each statistical procedure.

Data analysis and exploration.   We agree wholeheartedly with the trend of including more exploratory and confirmatory data analysis in statistics courses and have incorporated an extensive amount into the text and exercises. We also recognize, however, that not all readers will have access to computers and therefore have presented data analysis in a way that does not require using a computer even though one is recommended.

Detailed and careful explanations.   We have included every step of explanation we think a typical reader might need. Our guiding principle is to avoid "cognitive jumps" and thereby to make the learning process smooth and enjoyable. We believe detailed and careful explanations result in better understanding.

Data sets. In most examples and exercises, we have presented raw data in addition to summary statistics. Including raw data gives a more realistic view of statistics and provides an opportunity for students to solve the problems by computer if they wish.

Procedure boxes. To help the reader learn to apply statistical procedures, we have developed easy-to-follow, step-by-step methods for carrying them out. For ease in locating, each procedure is displayed with a color background. A unique feature of this book is that when a procedure is illustrated by an example, each step in the procedure is presented again within the example. This repetition serves a twofold purpose: It shows how the procedure is applied and helps the reader master the steps in it.

Procedure index. Given the numerous statistical procedures, it is sometimes difficult to find a specific one, especially when the book is being used for reference. To alleviate this problem, we have included a *procedure index* on the back inside cover of the book, providing a quick and easy way to find the required procedure for performing any particular statistical analysis.

Computer simulations. We have incorporated many computer simulations in both the text and the exercises. These simulations serve as pedagogical aids for understanding complex concepts such as sampling distributions. Readers can benefit from this material even if they do not have access to a computer.

Computer usage. Today, almost all professional applications of statistics are done by computer. It is therefore important that every student of statistics have at least some familiarity with statistical software. Although we have chosen Minitab[†] to illustrate the use of statistical software, the text has been written so that the instructor is free to select other statistical software packages. The Minitab discussion in this book is self-contained.[‡] All computer material is *optional,* but recommended.

The computer sections are integrated as optional subsections occurring immediately following the particular statistical concept under consideration. In each such subsection, we explain how Minitab can be used to solve problems that were solved by hand earlier in the section. Each solution consists of introducing the required commands, displaying the computer output, and interpreting the results.

Additionally, computer exercises (clearly marked) are incorporated into the exercise sets. Three types of computer exercises have been included. The first type asks the reader to interpret computer printouts; no knowledge of or access to statistical software is necessary for these exercises. The second type asks the reader to use Minitab or some other statistical software to solve exercises that were presented previously for hand solution; all Minitab commands required for these computer exercises will have already been discussed in the text. The third type of computer exercise asks the reader to use statistical software to perform a computer simulation; these exercises are designed to provide concrete illustrations of some of the more complex concepts (e.g., sampling distributions) and to demonstrate how the computer can be used to reveal statistical facts.

[†] Minitab is a registered trademark of Minitab Inc., 3081 Enterprise Drive, State College, PA 16801. Telephone: 814-238-3280. Fax: 814-238-4383. We would like to thank Minitab Inc. for their assistance.

[‡] Additional details and further topics, such as writing and using macros, are provided in the *Minitab Supplement* to this book.

For maximum flexibility we have allowed for three options in computer coverage: (1) no coverage, (2) coverage omitting the Minitab-command discussions but including the interpretation of computer printouts, and (3) coverage of Minitab commands (or those for some other statistical software) and the interpretation of computer printouts.

Choice of session-command or menu-command interface.   Traditionally, Minitab has been applied by typing (session) commands at the MTB > prompt. But in response to the strong emergence of graphical user interface (GUI), Minitab has recently introduced a menu interface as well, so that commands can also be executed by choosing them from menus and completing dialogue boxes. For each application of Minitab, we explain separately both the required session commands and the required menu commands. It is not necessary to cover both session commands and menu commands (although experienced users of Minitab often intersperse them).

Biographical sketches.   Each chapter begins with a brief biography of a famous statistician. Besides being of general interest, these biographies help the reader obtain a perspective on how the science of statistics developed.

Chapter introductions and chapter outlines.   Also included at the beginning of each chapter is a description of the chapter contents and an explanation of how the chapter material relates to the text as a whole. As a further aid, a chapter outline, which follows the chapter introduction, lists the major sections in the chapter.

Case studies.   Case studies range from the classic to the current. At the beginning of each chapter, a case study is described briefly; at the end of the chapter, when the reader has studied the relevant concepts, the case study is considered in detail. Exercises are provided for all case studies.

Definitions, formulas, key facts.   As an aid to learning and for reference, we have prominently displayed all definitions, formulas, and key facts. These items are enclosed by a color rectangle to make them easy to locate.

Real-life examples.   Because we believe that the majority of students learn by example, every concept discussed in the book is illustrated by at least one detailed example. The examples are, for the most part, based on real-life situations and have been chosen for their interest as well as their illustrative value.

Extensive and diverse exercise sets.   We have constructed exercise sets that are both extensive and diverse. Most of the exercises are based on information found in newspapers, magazines, statistical abstracts, and journal articles; sources are explicitly cited. The exercises are designed not only to help the reader learn the material, but also to show that statistics is a lively and relevant discipline.

Since students in elementary statistics courses often have different mathematical backgrounds, we have included three levels of exercises: basic, intermediate, and advanced. Every exercise set contains several basic exercises. The *basic exercises* provide applications of material presented in the text, and every reader should master these. We have organized the basic exercises so that each concept is covered by at least two problems; for each odd-numbered basic exercise that involves a particular concept, there is also an even-numbered basic exercise that involves that same concept. A single color bullet (·) preceding an exercise number identifies the exercise as a basic one. The answers to the odd-numbered basic

exercises are presented in Appendix C; the answers to the even-numbered basic exercises are in the *Instructor's Solutions Manual.*

Most exercise sets also include intermediate and advanced problems. The *intermediate exercises* contain supplementary material that is not necessarily covered in the text but that may interest some of the more highly motivated students. Two color bullets (· ·) preceding an exercise number identify the exercise as an intermediate one. The *advanced exercises* cover abstract concepts, theory, and algebraic derivations. Those exercises are intended for students with special mathematical background and aptitude. Three color bullets (· · ·) preceding an exercise number identify the exercise as an advanced one. The solutions to all intermediate and advanced exercises are in the *Instructor's Solutions Manual.*

Chapter reviews. Frequently, students in elementary statistics courses feel a certain amount of anxiety and confusion about how to study and review. To help the student, a chapter-review section appears at the end of every chapter. The chapter reviews include (1) a list of key terms with page references, (2) formulas with page references, (3) chapter objectives, and (4) a review test. These pedagogical aids provide the student with an organized method for reviewing and studying. The answers to the review tests are given in Appendix C.

Database exercises. Appendix B contains a printout of a database obtained by randomly selecting 500 Arizona State University sophomores. Seven variables are considered for each student: sex, high-school GPA, SAT math score, cumulative GPA, SAT verbal score, age, and total hours. At the end of each chapter review, a section entitled "Using the Focus Database" presents several exercises about the database. These exercises are optional and are to be done by computer. We assume that the student knows or will be taught any additional commands required to carry out these exercises.

Formula/table card. A detachable formula/table card is provided with the book. This card contains all of the formulas and many of the tables that appear in *Elementary Statistics.* The formula/table card is useful for quick-reference purposes; many instructors also find it convenient for use with examinations.
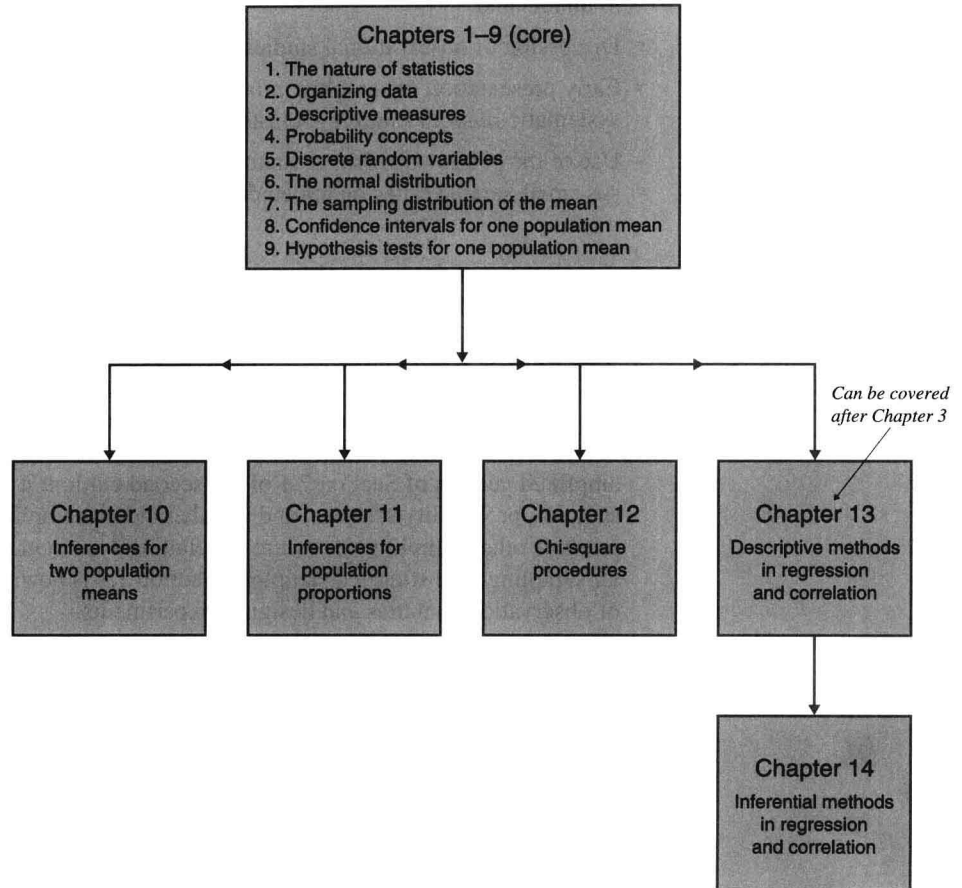
Organization

The text offers a great deal of flexibility in choosing material to cover. Chapter 1 introduces the nature of statistics, sampling designs, observational studies, and designed experiments; an optional introduction to Minitab is presented as well. Chapters 2 and 3 provide the fundamentals of descriptive statistics and exploratory data analysis.

Chapters 4–6 present probability, discrete random variables, and the normal distribution. Chapter 7 introduces the concept of sampling distributions and provides a detailed discussion of the sampling distribution of the mean. Following that, Chapters 8 and 9 examine confidence intervals and hypothesis tests for one population mean. We consider Chapters 1–9 the core of an elementary statistics course.

Chapter 10 presents inferences for two population means, Chapter 11 inferences for population proportions, and Chapter 12 chi-square procedures (goodness-of-fit test, independence test, and inferences for a population standard deviation).

We have divided the traditional material on regression and correlation into two chapters. Chapter 13 examines descriptive methods in regression and correlation and can be covered at any time after Chapter 3. Chapter 14 presents inferential methods in regression and correlation and can be covered once Chapters 9 and 13 have been completed.

The following flowchart summarizes the preceding discussion and shows the interdependence among chapters. In the flowchart the prerequisites for a given chapter consist of all chapters having a path leading to that chapter.

Chapters 1–9 (core)
1. The nature of statistics
2. Organizing data
3. Descriptive measures
4. Probability concepts
5. Discrete random variables
6. The normal distribution
7. The sampling distribution of the mean
8. Confidence intervals for one population mean
9. Hypothesis tests for one population mean

Can be covered after Chapter 3

Chapter 10
Inferences for two population means

Chapter 11
Inferences for population proportions

Chapter 12
Chi-square procedures

Chapter 13
Descriptive methods in regression and correlation

Chapter 14
Inferential methods in regression and correlation

Changes in the Third Edition

We have made significant changes in the third edition of *Elementary Statistics*. All chapters have been rewritten for the purpose of updating and expanding the material, fine-tuning the organization, and adding new sections where appropriate. The entire manuscript has been edited to provide a smoother and more succinct presentation.

Among the global enhancements to the third edition of *Elementary Statistics* are:

- Extensive coverage of data analysis.
- Increased interpretation of concepts in text and exercises.
- Expanded use of graphics to illustrate and explain concepts.
- Significant increase in the number and type of exercises.
- Fully expanded treatment of P-values with all hypothesis tests.
- Extensive computer simulation and resampling to make complex concepts easier to understand.
- Discussion of observational studies and designed experiments.
- Early presentation of sampling procedures, including simple random sampling, systematic random sampling, cluster sampling, and stratified sampling.
- Use of the σ-known versus σ-unknown criterion instead of the large-sample versus small-sample criterion to decide which inferential procedures to employ for means.

A complete list of the local enhancements to the third edition of *Elementary Statistics* is too long to present here; but a list of some of the most important ones follows.

- Chapter 1: The optional section on Using the Computer (Section 1.4) now includes a discussion of the SET, READ, and NAME commands to provide an alternative to the Data Editor for data entry and column naming. Section 1.5 is new; it discusses how to decide whether a study is necessary. Section 1.6 is a revised and amplified version of Section 7.1 of the second edition; it introduces representative samples, probability samples, and simple random sampling. Section 1.7 is new; it presents other sampling procedures, including systematic random sampling, cluster sampling, and stratified sampling. Section 1.8 is new; it contains a discussion of observational studies and designed experiments.

- Chapter 2: Section 2.1 has been extensively revised; it now discusses variables (as well as data) and uses a simpler and more modern classification scheme. Section 2.5 is a new section devoted to interpreting and understanding distribution shapes.

- Chapter 3: This chapter has been expanded to include more data analysis. The major change is to Section 3.6, now entitled "The Five-Number Summary; Boxplots." Contained in that section is a thorough examination of outliers, a general definition of quartiles, a presentation of the interquartile range, and a discussion of modified boxplots (in addition to ordinary boxplots).

- Chapter 5: In Section 5.1 a new subsection presents an interpretation of probability distributions by comparing the relative-frequency distribution of a simulation of a random variable with the random variable's probability distribution. The optional computer subsection that follows shows how Minitab can be used to perform the simulation (using the RANDOM command and DISCRETE subcommand). In Section 5.2 the material on interpreting the mean of a random variable has been augmented by two computer simulations and associated graphical displays. In

Section 5.4 binomial distributions are classified according to shape (i.e., skewness or symmetry) as an aid for visualizing binomial distributions in general and understanding the normal approximation to the binomial distribution in particular.

- Chapter 6: A cumulative normal table (areas between $-\infty$ and z) is now used. This table is simpler to learn and apply than the normal table giving areas between 0 and z. Moreover, the cumulative normal table is consistent with Minitab's CDF and INVCDF commands (which are presented in the optional computer subsection in Section 6.1). To give the reader a better understanding of the normal distribution, a simulation of one is presented and compared to the theoretical distribution. This simulation is featured in Section 6.4 (Example 6.20). The optional computer subsection at the end of Section 6.4 shows how Minitab can be used to perform the simulation (using the RANDOM command and NORMAL subcommand). A new section on normal probability plots appears as Section 6.5. In Section 6.6 probability histograms and the concept of distribution shape are now used to help explain the normal approximation to the binomial distribution.

- Chapter 7: The section on sampling, found in Section 7.1 of the second edition, has been expanded and moved to Chapter 1 (see above).

- General changes to Chapters 8 and 9: These two chapters have been extensively revised. They have been rewritten to include the use of exploratory data analysis as a preliminary to inferential analysis. Furthermore, inferences for one population mean are treated differently than in previous editions. Instead of employing the large-sample versus small-sample criterion for deciding which procedure to use, the σ-known versus σ-unknown criterion is employed.

- Other changes to Chapter 8: In Section 8.3 we have substituted the term *margin of error* for *maximum error of the estimate* since the former term is more intuitive and is usually the one found in newspapers and magazines. Section 8.4, which covers the one-sample t-interval procedure, also includes a discussion on how to choose the correct confidence-interval procedure.

- Other changes to Chapter 9: The difference between statistical significance and practical significance is explained in Section 9.3, as is the relationship between hypothesis tests and confidence intervals. In Section 9.4 the method for using the P-value to assess the evidence against the null hypothesis has been improved. Section 9.5, which covers the one-sample t-test, also explains how to estimate the P-value of a t-test using the t-table; additionally, this section discusses a strategy for choosing the correct hypothesis-testing procedure.

- Chapter 10: Section 10.1 now discusses only the sampling distribution of the difference between two means for independent samples. Procedures based on known standard deviations (two-sample z-interval procedures) have been relegated to the exercises, because population standard deviations are rarely known. Section 10.3 is new; it examines nonpooled t-interval procedures and compares the pooled and nonpooled methods. Exploratory data analysis has been incorporated throughout the entire chapter as a method for helping select the correct inferential procedure.

- Chapter 11: This chapter is new; it includes the material on inferential methods for one and two proportions found in Sections 8.5, 9.6, and 10.5 of the second edition. Margin of error and sample-size determination for one proportion are now

discussed in Section 11.1. Margin of error and sample-size determination for two proportions are now discussed in Section 11.3. Computer simulations have been incorporated; these appear both in the text and in the exercises.

- Chapter 12: A discussion of how to employ the P-value approach to perform chi-square procedures has been included; this coverage is in addition to the classical approach using critical values. Segmented bar graphs are presented to provide a visual interpretation of two characteristics of a population being statistically independent (nonassociated). The nonrobustness of chi-square inferences for a standard deviation is also discussed.

- Chapter 13: Outliers and influential observations are covered in Section 13.2. The terms *predictor variable* and *response variable* are explicitly introduced in Section 13.2.

- Chapter 14: Analysis of residuals for linear regression is presented in Section 14.1, thus providing a method for checking the assumptions for regression inferences.

Supplements

The following supplements have been prepared specifically to accompany the third edition of *Elementary Statistics*.

MINITAB$^®$ Supplement.   This supplement, written by Professor Peter W. Zehna, provides in-depth coverage of Minitab, augmenting that given in the book. It is designed to be used in conjunction with *Elementary Statistics* and is keyed to the book. No prerequisite knowledge of computers or statistical software is presumed.

Instructor's Solutions Manual.   This manual, prepared by Professor David Lund, contains detailed, worked-out solutions to all exercises in *Elementary Statistics*.

Student's Solutions Manual.   This manual, also written by Professor Lund, presents detailed, worked-out solutions to every fourth exercise in *Elementary Statistics*.

OmniTest III Computerized Testing.   This unique testing software offers an abundant supply of quizzes, tests, final examinations, and additional instructional exercises. Features include multiple-exam versions, customized question editing, on-screen preview and edit functions, and pull-down menus. *OmniTest III* is available in both IBM and Macintosh versions.

Printed Test Bank.   This supplement provides several printed examinations for each chapter of *Elementary Statistics*.

DataDisk.   This floppy disk, available in both IBM and Macintosh versions, contains files for the Focus database and for the data sets appearing in the case studies and odd-numbered basic exercises in *Elementary Statistics*. DataDisk makes it possible to store and use these data sets without having to enter them manually.

Transparency Masters.   Many of the text's figures, tables, and procedure boxes have been reproduced on transparency masters for classroom use.

Acknowledgments

It is our pleasure to thank the following reviewers, whose comments and suggestions most certainly improved the third edition of *Elementary Statistics:*

Scott Balcomb, Saint Joseph's College

David Lund, University of Wisconsin, Eau Claire

Donald Mason, Elmhurst College

Theodore Moore, Mohawk Valley Community College

Reed Parr, Salt Lake Community College

Bryan Stewart, Tarrant County Junior College, Northwest Campus

Lav Vrdoljak, Barry University

Ruth Yarar, Metropolitan State College of Denver

We also wish to thank Professor Michael Driscoll for helping us select the statisticians for the biographical sketches; and Professors Charles Kaufman, Sharon Lohr, Bernard Morzuch, and Dennis Young, with whom we had several illuminating discussions. Our special thanks go to Professors Matthew Hassett and Ronald Jacobowitz for their many helpful comments and suggestions.

To Professor Larry Griffey, we express our appreciation for his formula/table card. We are grateful to Professor Peter Zehna for preparing the *Minitab Supplement* and for many interesting and informative conversations. Our thanks go as well to Professor David Lund for writing the *Instructor's Solutions Manual* and *Student's Solutions Manual.*

We are grateful to Dr. George McManus and Mr. Gregory Weiss for supplying the data from their study of zooplankton nutrition in the Gulf of Mexico. We also extend our appreciation to Professor John Tukey for taking the time to provide us with autobiographical information; to Maureen Quinn of The Gallup Organization for supplying the information used in the case study of Chapter 12; to Cathy Akritas and Rob Kelly of Minitab Technical Services; to Pam White and Susan Shirey of the Minitab Author's Assistance Program; to Jennifer Wang for obtaining and updating much of the data used in the examples and exercises in the book; and to Mary Neary for furnishing the Focus database.

Our thanks to our text designer, Rebecca Lemna; to our cover designer, Marshall Henrichs; to our proofreader, Phyllis Coyne et al.; to our PostScript consultant, Arthur Ogawa; and to our prepress buyer Caroline Fell Skurat. Thanks also to Berthold Horn, Louis Vosloo, and Jordan Backler at Y&Y, Inc.; to George Morris, Brian Morris, and the others at Scientific Illustrators; and to John Smiroldo at Pure Imaging.

Special thanks to our editor, Julia Berrisford; to our editorial assistant, Kate Morgan; to our production supervisor, Patsy DuMoulin; to our copy editor, Sally Stickney; and to our marketing manager, Michelle Babinec.

Finally, we thank Carol Weiss for her participation in almost every aspect of development and production, for researching and writing the biographies, and for typesetting the book using the TEX typesetting system.

Tempe, Arizona   *N.A.W.*

Contents

. . . . .