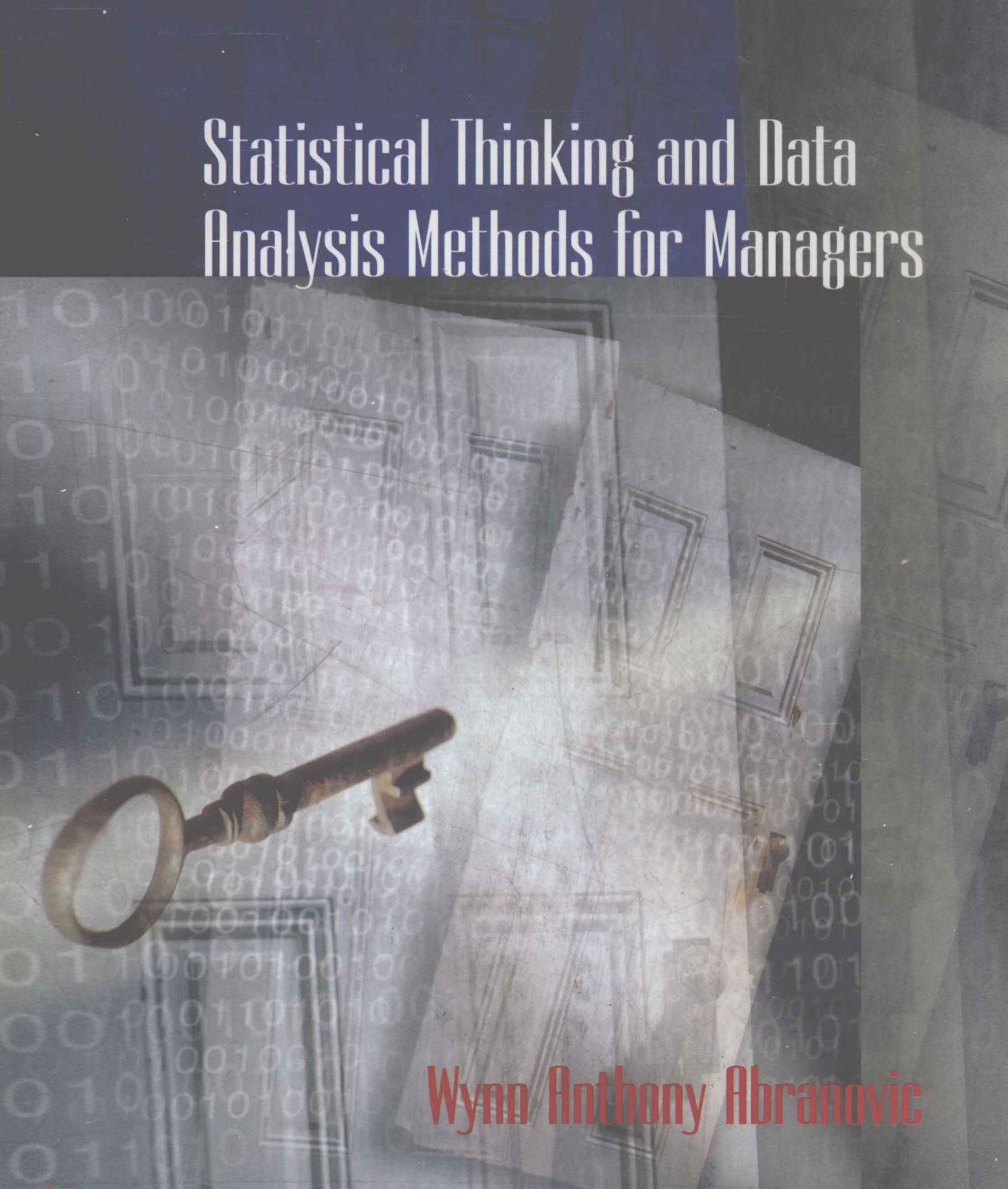
Statistical Thinking and Data Analysis Methods for Managers



Wynn Anthony Abranovic

Statistical Thinking and Data Analysis Methods for Managers

Wynn Anthony Abranovic
University of Massachusetts at Amherst

ADDISON-WESLEY

An imprint of Addison Wesley Longman, Inc.

Reading, Massachusetts • Menlo Park, California • New York • Harlow, England
Don Mills, Ontario • Sydney • Mexico City • Madrid • Amsterdam

Sponsoring Editor: Michael Roche
Developmental Editor: Maxine Effenson Chuck
Project Coordination: Elm Street Publishing Services, Inc.
Art Development Editor: Vita Jay
Design Administrator: Jess Schaal
Text and Cover Design: Jeanne Calabrese
Cover Illustration: Kevin Irby
Production Administrator: Randee Wire
Compositor: Interactive Composition Corporation
Printer and Binder: R. R. Donnelley & Sons Company
Cover Printer: Phoenix Color Corp.

Statistical Thinking and Data Analysis Methods for Managers

Library of Congress Cataloging-in-Publication Data

Abranovic, Wynn Anthony.
Statistical thinking and data analysis methods for managers / Wynn Anthony Abranovic. -- 1st ed.
p. cm.
Includes index.
ISBN 0-673-99296-9
1. Industrial management--Statistical methods. 2. Industrial management--Statistical methods--Data processing. 3. Minitab for Windows. I. Title.
HD30.215.A27 1997
519.5--dc20 96-7505

1 2 3 4 5 6 7 8 9 10—DOC—00 99 98 97 96

Dedicated to the Curious Student

PREFACE

To The Reader

In this book you will be introduced to statistical knowhow that will help you grapple with interesting problems. For example, when Toyota, General Motors, Ford, or Chrysler aspires to produce automobiles of world-class quality, those of you who have had your car repaired recently or who are considering a purchase might be interested in company quality reputations. You might be interested in survey results about owner satisfaction by J.D. Power and Associates. You might be interested in how experiments are used to improve product designs. Quality improvements in product designs, manufacturing processes, and final inspections frequently depend on companywide understanding and application of statistical methods. Other types of examples where statistical methods play an important role are numerous and apparent from simply reading the daily newspaper.

By reading this book, you will learn how to use statistical methods in a variety of applications. Statistical skills will help you to quickly gain experience in unfamiliar work environments and to survive in today's competitive workplace. You will learn why statistical methods work and what their limitations are.

What Is Statistics, and Why Is It Challenging?

Statistics is the art and science of interpreting numerical observations. The analysis of numerical information helps us think about answering interesting questions, such as those on product quality.

The challenge is that statistics, like tennis, basketball, golf, baseball, medicine, economics, and management, can't be learned in a day. You have to take the time to understand the components of the statistics game, the art of putting the components together, and how to keep your game in tune through practice.

Modern statistics books are likely to be weighty because they include so many methods. Yet, they contain considerable and valuable knowledge. The sheer multitude of statistical methods makes the subject both difficult and interesting for the novice as well as the expert.

Nevertheless, most students of management, economics, the social sciences, and other disciplines are expected to learn the fundamentals of statistics in one or two semesters. What is expected of the novice is often, simply, too much. Such a stiuation can create anxiety for both student and teacher. Improving this situation is the reason for this book.

Modern Textbook Design Themes

While few claim to have found all the solutions to the teaching and learning problems in introductory statistics, many have tried, as evidenced by the proliferation of introductory textbooks. The variety of texts alone suggests that teachers work hard at delivering a unique, high-quality classroom experience for the student. Several common features that support the efforts of teachers and students have emerged in textbook designs. This textbook was written with all the following features in mind:

1. Applications and methodology are integrated to give students reasons to learn the basic methodology. Applications that students already know something about are used, such as the popcorn experiment in Chapter 1, where brand and amount are related to yield.

2. The text was written for the nonmathematician. It drains the technical swamp where busy students can get bogged down and discouraged, without sacrificing statistically rigorous thinking. For example, in regression analysis, statistical software can be used to ease the formula-based computational burden.

3. Examples and exercises are relevant and plentiful, using mostly real data and no less than realistic data. Data sets on diskettes are included for convenient computer access. For example, for students of management, such diverse fields as personnel, quality control, economics, marketing, accounting, operations management, and finance are drawn on.

4. Some case studies are provided, usually as examples or exercises, to frame the role that the analysis plays in the big picture as well as to motivate and provide a discussion platform. For example, one case is about a company that machines mold cavities out of steel for injection molding of plastic parts. The company was seesawing from profit to loss and needed to know why it was losing so much money on some jobs and how to cut costs. The data analysis results led to managing large jobs differently from small ones. The profit picture improved and operations ran smoother.

5. Computer output is shown as well as chapter appendixes of MINITAB commands. Therefore, students can walk through the chapter examples actively on their own. Those who prefer other software, such as SAS, SPSS, or BMDP, can adapt easily. The option to use the computer is available in 18 chapters, with little investment needed by the instructor. The software encourages self-study and reduces the teaching time allocated to computational formulas. More time can be spent on conceptual interpretation rather than computation. Nevertheless, standard computational formulas can be demonstrated when there is a need to look under the hood at the engine.

6. Recognizing modern statistical practice, we introduce exploratory data analysis (EDA) tools such as stem-and-leaf diagrams and boxplots. However, they do not overshadow tradtional methods. Traditional foundations of statistical inference are integrated with data analysis ideas and take advantage of the visual (graphical) flavor of EDA.

7. Quality control concepts are included because global industrial competition requires it. Furthermore, quality control's high corporate profile offers proof that students must have statistics in their portfolio of skills.

8. To simplify and reveal the connecting thread that runs through statistics, a kernel of building block concepts is provided for organizing the textbook, thus smoothing the transition from simple ideas to complex ones. Where possible, graphical displays are used to build visual representations of methodology, letting each of the data speak for itself. For example, a pair of side-by-side boxplots can test, informally, for two equal population medians, ideawise leading to tests for population means. Several side-by-side boxplots can illustrate regression analysis concepts.

9. This textbook is user friendly and very flexible. It accommodates different teaching styles, variations in course content, and a variety of learning styles. For example, an instructor who uses a computer lab intensively (where students work through chapter appendixes) can cover two semesters of statistics with few lectures. On the other hand, a lecture-intensive course can be supplemented with a few computer lab sessions.

10. The writing style kindles student interest and makes teaching easy and effective. The reader relates to the subject matter because it covers a broad range of real-life scenarios, including managerial issues, social events, business decisions, sports, economics, and other topics familiar to the reader.

How Does This Textbook Stand Out from the Crowd?

Many statistics books draw on at least some of the features listed above, but they often try to be all things to all people, thus losing their identity and becoming clone-like in appearance. This textbook preserves its identity with the following pedagogical features:

1. *Writing Style.* Students tell me this book is fun to read because it's written in a way that is clear, down to earth, easy to understand, and motivating. For example, Chapter 1 begins with an observation on a December day that raises the question, Why are there still ducks on the campus pond? To answer the query, we suggest that it might help to collect data on where ducks winter and see what the analysis reveals.

The chapters begin with short motivational scenarios that give students a reason to read on. Chapter 3 introduces a statistical model in which the mean is a typical value and an observation's deviation from the mean is an error. Because these concepts can be important in executive decision making, Chapter 3 begins like this:

> Stories about major executive decisions and costly gambles are plentiful. Consider Barnaby Feder's account of General Electric's high-stakes locomotive wager.
>
> GE embarked on a $500 million modernization plan that got derailed as diesel orders slumped. For GE, the sales forecasts were wrong. Typically, domestic railroads buy 1,700 locomotives a year, but only 300 new locomotives were bought in the year that the modern facilities were available. This forecasting error reflected management's failure to understand how much the actual demand might deviate from the typical demand. This error cost GE a lot of money.

2. *Concept Linking.* Seldom is a concept presented and not used shortly thereafter. Many concepts are tied to graphical building blocks such as histograms, scatterplots, stem-and-leaf diagrams, boxplots, dotplots, and the like. As an example, when introducing hypothesis tests for the difference between two means, side-by-side boxplots may lead to tentative, graphically based conclusions, which are confirmed by more formal tests that follow. Such visual, intuitive lead-ins to formal methods, common in this text, help students learn new ideas by fitting them into a familiar context and provide concept repetition by reusing what was learned earlier.

3. *Concept Distribution.* New concepts are presented in bite-size chunks. Thus, high concept density is avoided when introducing complex ideas. Students are not intimidated by new methods, techniques, or displays.

4. *Flexibility.* Primarily, this text is designed for a one- or two-semester introductory course without prerequisites. The instructor can vary the speed of coverage and, with the addition of computer lab sessions and exercises, use the text at the undergraduate or graduate levels, with MBAs in mind at the graduate level.

5. *Artful Use of Simulation.* A major use of modern computers is for "what if" simulations. Many examples and exercises offer students and instructors the opportunity to actively experiment on the computer and summarize the results statistically. The simulation approach augments the multitude of real-data examples and exercises in the text.

6. *Examples.* To give you a flavor of the examples, below are some listed in sequence for Chapter 1.

> Chapter 1 *The Kaleidoscopic World of Applied Statistics:* The Scurvy Experiment; Box and Hunter's Popcorn Experiment; The Strawberry

and Williams Observational Study (Baseball); Automobile Inspection Observational Study; The Gypsy Moth Population; Prediction—Maine Lobster Outlook: Down; Key Variables As Early Warning Signals for the Economy; Why Aren't People Buying American Cars; Tom Martin Creek.

7. *Optional Computer Use for Addtional Flexibility.* Where statistical software is used in examples, a chapter appendix is provided of MINITAB commands so the student can do each example as well as similar exercises. Each appendix listing of MINITAB commands includes line-by-line commentary, explaining what the commands do. The students can, in effect, go to the computer lab and work through examples in the entire text on his or her own, roughly completing a chapter a week.

Such guided, walk-through-the-examples type of activity produces the usual, barren computer output. MINITAB prints or creates a file of this output. The file can be retrieved by a word processor. This document can be edited and integrated into a formal report, in which the student adds commentary. The comments, which parrot the textbook discussion, assure the instructor that the students are reading and mastering the key ideas. Extra exercises can be assigned and done without such structured guidance. Often, deliverable products from these activities can be substituted for examinations as the students engage in active learning.

Alternatively, the file of MINITAB results can be printed in its original barren form and the students can add commentary by hand. This reduces lab time because the word processor is not needed and comments can be written at home.

In addition to data sets on diskettes, the MINITAB appendixes are available in the form of chapter macros. A macro command produces output for a whole chapter of examples. Therefore, what the students do, say, for an hour-and-a-half, one command at a time, can be done in a minute by a macro. Under instructor control, the macro feature can be useful when the instructor feels the students have gained enough command familiarity, proceeding one command at a time. Also, the output from these macros can be projected on a screen for lecture purposes. Macros can speed up the course, they relieve tedium at the right time, they can be used by teaching assistants, and they can form a basis for team projects, exams, and presentations by students.

Of course, all this works best if you are using the latest version of MINITAB in a lab of networked personal computers. However, the chapter appendixes are detailed enough so that one experienced in SAS, SPSS, BMDP, or some other software system can substitute appropriate commands.

Although this text can be used in a regular class situation without computer laboratory work, I find it easy and effective to conduct most of my classes in a computer lab. Typically, I give a few start-up lectures and then move the class to the computer lab, fitting in occasional review lectures when necessary. After the start-up phase, I sit back and let the course run by itself.

Students refer to this phase of the course as "being on cruise control." Avoiding the lecture format, I find it quite pleasant to get paid for the less strenuous task of being a helper, a coach, and having fun teaching. Students become gladiators rather than spectators.

One Course or Two?

Our texbook design is flexible to accommodate the wide variety of feasible course designs needed for meeting local teaching requirements. Depending on the depth and speed of coverage, the first 12 chapters form a sound basis for a one- or two-semester introductory statistics course. The remaining chapters provide additional topics for higher-level courses, particularly when the students are exposed to two full semesters of statistics.

You will notice that the chapter sequencing is similar to what has become the traditional pattern. However, as in the practice of applied statistics, exploratory data analysis takes on a more important role than in the past, and EDA is used as a lead-in to traditional topics with emphasis on its graphical portrayal of concepts. The instructor has the option of emphasizing the classical analysis, the exploratory analysis, or a balanced blend of the two approaches.

As the experienced instructor might expect, fitting the first 12 chapters into a single semester is a squeeze play, possibly with something thrown out. Yet, in two semesters, it looks like a walk.

Some of This Textbook's Design Differences

This text is designed to be a little different from certain chapters in other texts. The following items alert you to some of these differences.

1. After a lively and broad introduction in Chapter 1, Chapter 2 concentrates on descriptive exploratory methods for actively getting familiar with data sets. Such methods require few assumptions about the data. It is not until Chapter 3 that the classical measures of central tendency and variability are introduced. Chapter 4 on probability theory and Chapter 5 on the normal distribution follow to enrich the framework for thinking about observed data.

2. Chapter 6 is integrative and makes use of methods from the preceding chapters to make data sets, such as the gross domestic product (GDP), more meaningful to students and managers. The, new material on sampling distributions, confidence intervals, and hypothesis testing is presented in Chapter 7.

3. Exploratory ways of comparing and relating two variables are covered in Chapter 8. The front end of Chapter 9 discusses planning and designing experiments prior to the simplest kind of an experiment that formally compares sample means. Furthermore, it is shown that sample means can be compared using the usual t tests. However, similar results are obtained from simple regression analysis. Thus, these methods are introduced and related in the same context. More advanced applications of these methods that come later are easier to understand because of this presentation.

Many introductory texts don't say much about planning experiments. If they do, it is delayed until more complicated designs are presented in an analysis of variance chapter.

4. Chapter 10 consolidates methods for categorical variables. For example, yes or no responses on questionnaires lead to confidence intervals and hypothesis tests on proportions and differences in proportions. Other texts discuss these intervals and tests in parts of different chapters. Thus, the methods appear scattered about in many books. The remainder of Chapter 10 is traditional. Some instructors may prefer to cover Chapter 10 later, perhaps after Chapter 12.

5. A careful development of the simple regression model and theory occurs in the beginning of Chapter 11. Some instructors may give this beginning section less emphasis than others. When less emphasis is given, the instructor can concentrate on the applications that follow the theory.

Chapter 11 presents the pooled t test, simple regression, and one-way analysis of variance as alternative ways of doing the same analysis. Most texts discuss the pooled t test earlier and, unfortunately, never relate it to these other methods.

6. Because the theory for multiple regression analysis is formidable, Chapter 12 begins multiple regression analysis by example and as an extension of simple regression. This is followed by a technical reference section on the theory. Thus, instructors can use a very applied approach or one that is more theoretical. This makes Chapter 12 long, but most instructors will not find it necessary to cover it all.

Some data sets are not listed in the exercises at the end of Chapter 12. Rather, they are available as files on a data disk.

7. Many chapters beyond Chapter 12 include special topics that can be used to enrich earlier chapters. For example, Chapter 13 covers special probability distributions. It extends the probability theory of Chapter 4 and goes well beyond the normal distribution in Chapter 5. Even less common distributions are mentioned. Cumulative distribution functions and their inverses are covered thoroughly.

Instructors who want their students to have a larger dose of probability theory would cover much of Chapter 13 and perhaps, include Chapter 14. Chapter 14, on statistical quality control, demonstrates important uses of the distribution theory from Chapters 4, 5, and 13.

In Chapter 18, the nonparametric methods are rough and ready and easy to compute. As a result, these methods are versatile and good for taking a first look at a data set. Therefore, with these methods some instructors may supplement subject matter in Chapters 2, 7, 9, and 15.

As you can see, the special topics chapters are organized and located in proximity to related chapters where possible. Furthermore, the topical coverage exceeds what most instructors need, providing options and flexibility.

Chapter Pedagogy

All chapters have a similar architecture. Some common chapter pedagogical features are

Objectives A short list of key objectives is presented.

Prelude A broad overview of the chapter is given in a few short paragraphs.

Introductory Discussion An interesting, attention-getting, real-life example or situation is used to expand upon the chapter objectives, getting more specific than the prelude.

Sections Sections break the chapters into manageable chunks. Some sections are independent of others and may be treated as optional, giving the instructor some flexibility within a chapter. Sections develop specific topics. For example, Chapter 2 presents a number of exploratory tools and has the following section topics: The Stem-and-Leaf Diagram; Computer Results; A Primitive Plot; Summary Statistics, Medians, Quartiles and Extremes; Summary Statistics, Computer Results; Checking for Randomness to See If the Summary Statistics Are Valid; Boxplots; Stem-and-Leaf Diagram Variations; Exploring Several Variables, One at a Time; A Population, a Sample, Different Samples with Different Summary Values.

To illustrate teaching flexibility, an instructor wishing to deemphasize computer use and exploratory methods could easily omit section topics on computer results and stem-and-leaf diagram variations. The scope of sections is limited to fundamental tools and concepts. This avoids clutter and does not overwhelm the student with too many new ideas at one time.

Each section contains interesting, real-world examples with plenty of exercises at the end of the chapter. There are MINITAB commands in the appendix for most examples. These commands (also in the form of macros)

can be used as guides to work through the chapter examples for assignments or for computer laboratory sessions that substitute for some lectures.

Statistical Highlights After detailed procedures or after a sequence of new terms and ideas, what to know summaries are provided within sections to help the student consolidate and review what was learned.

Key Terms Key terms are set in bold within the chapters for emphasis and listed at the end of the chapter.

Numerous Exhibits The emphasis on learning from visual graphics is brought out in the many exhibits.

Chapter Summaries These are brief reviews of key concepts.

Exercises Exercises are numerous and varied. Some are long enough to integrate several chapter sections.

Ancillaries

Many will find this textbook self-contained, however, an *Instructor's Manual* and *Test Bank* are also available. Additionally, data sets and macros are provided on diskettes.

Wynn A. Abranovic

Acknowledgments

My students motivated me to write this book. It seemed that I had searched for a decade for a course design that had both the old and the new ideas, enough conceptual rigor, and, perhaps most of all, the ability to capture the interest of students with or without a technical background. Therefore, students who left my course saying "I never knew that statistics and data analysis could be so interesting" gave me the incentive to go on when there were many better things to do, like ski, play tennis, or fly fish for trout.

Here's to you Elias Halamandaris, Chris Stefanou, Carol Savoy, Two Hugs, and the like. Former students seem to make more of an impression on me than I do on them. Elias still sees that I get a beautiful calendar each year from the Bank of Greece, even though it has been decades since we had dinner in a Greek taverna. Chris has such a thirst for knowledge and adventure that he simply puts his whole family on a Greek freighter and takes an odyssey covering the wide waters that connect three continents, all in the

spirit of exploration. Simply talking to Carol for a couple of hours makes a university's homecoming weekend seem too short, and it is nice to hear that my data analysis course was a favorite. And, to be sure that there is a little California sunshine in this book, thanks Angie for your individuality and for being such an inquisitive student. Incidentally, Two Hugs squeeze a big university into something smaller and certainly warmer.

Many others influenced me. Some, though I know only through their writing, shaped my thinking tremendously: those such as John Tukey, Frederick Mosteller, and Harry Roberts. Also, John Wilkinson and Richard Carter taught some great courses that got me interested in probability and statistics when studying clashed with Saratoga Performing Arts concerts. Richard Carter's foundation of probability problems remains unrivaled. A few simplified adaptations of his problems appear as exercises.

Special thanks to Barbara Dixon and Steve Archer at Willamette University. Barbara worked on the book's first draft, which formed the basis for the monster that has emerged. Steve stressed the need for motivational material that gives the student a reason to read. It took me a long time to figure out how to accomplish what he said was needed.

Mike Royer, Richard Lindgren, Chad Johnson, and The Great Ganguli encouraged me to play good tennis and to try to write a great book. Doug Tucker encouraged me, too, and also demonstrated perfect skiing turns. Thanks to old friends Dede, Jean, and Patty, who are long ago and far away, and Peggy who is not. My parents, brothers, and sister made it possible to persist and to write this book in an interesting style that combines text and graphics. In the early stages, Louis Wigdor did some first-rate editorial work for me. Vesta Powers made her usual extra effort to get me quality first-draft copies that could be tested in class. Thanks to my brother Tony who treated me to a fly fishing trip in Vermont in June—a break I needed after writing the first six chapters. Thanks to Barbara and Willard Weeks.

I am also grateful for the advice, feedback, and suggestions of the many reviewers which, undoubtedly, made this a better book. The reviewers were:

Mary S. Alguire, *University of Arkansas*
Wallace R. Blishke, *University of Southern California*
Michael Broida, *Miami University*
Roger Champagne, *Hudson Valley Community College*
Sangit Chatterjee, *Northeastern University*
David A. Cohen, *Northern Arizona University*
Eugene A. Enneking, *Portland State University*
Rob Godby, *Laurentian University*
Anil Gulati, *Western New England College*
Ramadan Hemaida, *University of Southern Indiana*
Rodney G. Hurley, *Hillsborough Community College*
Carla Inclan, *Georgetown University*
George A. Johnson, *Idaho State University*

Dae S. Lee, *Kentucky State University*
Douglas A. Lind, *The University of Toledo*
Jerrold H. May, *University of Pittsburgh*
Charles E. McClean, *Old Dominion University*
Joe D. Megeath, *Metro State University*
Ruth K. Meyer, *St. Cloud State University*
Daniel Mihalko, *Western Michigan University*
R. Ramesh, *State University of New York at Buffalo*
Charlene Robert, *Louisiana State University*
Alan Roshwalb, *Georgetown University*
Patrick Thompson, *University of Florida*

All the folks associated with Addison-Wesley Publishers were wonderfully helpful. They include Michael Roche, Melissa Rosati, Arthur Pomponio, Maxine Effenson Chuck, developmental editor at B. Czar Productions Inc., the reviewers, and others who helped place this book in the hands and minds of students and teachers are much appreciated. Maxine was relentlessly excellent in shaping my seemingly crude early work. Similar praise goes to Julie Webber, Nancy Moudry, and other professionals at Elm Street Publishing Services.

We all hope that students will want to keep this book.

W. A. A.

CONTENTS