




experiments, are discussed in an early chapter (Chapter 3). In our view, this topic should be covered before plunging into the details of data analysis. After all, how the data are collected dictates how they should be analyzed. More importantly, the method of data collection affects the quality of the data, and no amount of statistical magic can salvage a poorly designed study. This chapter discusses the important basic notions of sources of biases in data and presents randomization and blocking/matching as devices to avoid some types of biases.

4. Another topic that is not covered in most other textbooks is modern computer-intensive methods; in particular those based on the ideas of resampling and bootstrap. These are covered briefly in the chapter on nonparametric statistics (Chapter 14, Section 14.6). We subscribe to the widespread view that these methods will become increasingly important in the analysis of data as they permit a user to break the traditional shackles of strong distributional assumptions, typically normality and independence.
5. Chapter summaries are provided for a quick review of the main concepts and results. Tabular summaries of the basic formulas for standard probability distributions and tests and confidence intervals for selected parameters are given in respective chapters.
6. Each statistical method is presented and followed by one or two illustrative examples. A mathematical derivation of the method is generally given at the end of the section, and is marked with an asterisk. All advanced sections are marked with asterisks, so that the more elementary courses may skip them without disrupting the flow of the material.
7. In many cases the same examples are developed in steps in each chapter, analyzing the data in more depth as additional techniques are introduced. Some examples are carried across chapters to illustrate the applications of different methods to the same data sets.
8. We have paid particular attention to writing interesting and educational exercises. As far as possible, real data sets with sufficient background information are used. Many of the data sets are taken from the book *A Handbook of Small Data Sets* by Hand et al.<sup>1</sup>, all references to this book are abbreviated as *Small Data Sets*. Most exercises guide a student through several steps, each step addressing a different or an additional aspect of analysis. The emphasis is on urging a student to think critically and to interpret the analysis, not just calculate numbers even if they are the “right” answers.
9. Exercises for each chapter are grouped together at the end, but are categorized by chapter sections. A few drill type exercises are given at the beginning of each section, followed by more challenging ones. Advanced exercises are given at the end. They typically involve simple mathematical proofs or generalizations of some results or more extensive data analyses requiring methods from different sections of the chapter. Many of these advanced exercises give students glimpses into topics beyond the scope of the present text. As such, they may be assigned to graduate students or to bright, motivated undergraduates to encourage them to pursue further

---

<sup>1</sup> D. J. Hand, F. Daly, A. D. Lunn, K. J. McConway, and E. Ostrowski (1993), *A Handbook of Small Data Sets*, London: Chapman & Hall, 1993.

studies in statistics. Answers to selected odd-numbered exercises are included at the end of the text.

10. Many exercises require the use of a computer; these exercises are marked by the symbol . It is recommended that some of the smaller data sets be analyzed by hand as this experience adds to the understanding of the methods. The diskette packaged in the back of the text contains data sets for all exercises.

## SUPPLEMENTARY MATERIALS

The following supplementary materials are available to facilitate the use of this book.

1. **The Instructor's Solution Manual:** The manual gives detailed solutions including mathematical proofs, derivations, and graphical plots for all exercises in the book. It may be requested from the publisher upon adoption of the book.
2. **Data Disk:** The diskette packaged in the back of the text contains data sets for all exercises saved as ASCII, MINITAB, and Microsoft Excel files. Users who prefer to obtain the data files in SAS or SPSS format may download them from the web site.
3. **Web site:** A web site is set up at <http://www.prenhall.com/tamhane> where the following items will be posted as they become available: (i) errata, (ii) supplementary text material, (iii) book reviews, (iv) data files for all exercises, and (v) links to other useful sources of data sets.
4. **Text and Student Version Software Packages:**
  - Statistics and Data Analysis: and SPSS 9.0 Student Version Integrated Package:*** A CD-Rom containing the SPSS 9.0 for Windows Student Version and the data files from the text may be purchased as a package with the text book for a small additional charge.
  - Statistics and Data Analysis: and Minitab 12.0 Student Edition Integrated Package:*** A CD-Rom containing the Minitab Release 12.0 Student Edition and the data files from the text may be purchased as a package with the text book for a small additional charge.

## BACKGROUND AND THE TOOLS NEEDED

Mathematics and probability are the two important tools needed in the study of statistics. Although calculus is used relatively in modicum in this text, a certain mathematical maturity and ability that comes with a calculus background is required. In addition, basic knowledge of matrix algebra is necessary in the chapter on multiple regression. The chapter on probability (Chapter 2) covers the essentials needed to follow the remainder of the text. This chapter is provided for review and reference purposes for those students who have had a previous course in probability. This chapter may be used as a condensed introduction to probability in a combined probability-statistics course.

Another indispensable tool required in modern statistics is a computer. It is useful not just for the tedious task of number crunching, but is equally important for plotting informative graphics in a fast and interactive manner. We assume that a student has access to a desktop computer and a statistical package or a spreadsheet package with statistical capabilities. There are many excellent statistical packages to choose from, MINITAB<sup>®</sup>, SYSTAT<sup>®</sup>, SAS<sup>®</sup>, STATGRAPHICS<sup>®</sup>, S-PLUS<sup>®</sup>, and SPSS<sup>®</sup> to name a few. The Microsoft Excel<sup>®</sup> spreadsheet package can perform many statistical

tasks. It is not the purpose of this text to provide instruction in the use of any of these packages. After all, packages keep evolving, so any software instruction is likely to become obsolete even before the book is published. Fortunately, most of these packages now allow the user to perform complex statistical tasks at the click of a mouse. Also, each package has its strengths and weaknesses, and it is up to the individual instructor to choose the package that best suits the needs of his or her students. We illustrate examples using outputs from several different packages, but mainly from MINITAB, as it is one of the easiest to use.

### POSSIBLE COURSE SYLLABI

There is more than enough material in this book for a year-long course in statistics alone (with a probability course as a prerequisite). We realize that most curricula do not have the luxury of such an extensive course. At Northwestern University, we teach a two-quarter undergraduate sequence in statistics (with a quarter-long probability course as a prerequisite), which is taken mainly by industrial engineering (IE) majors with a sprinkling of other majors. For the benefit of IE majors we focus on quality control, reliability, and some business applications primarily through examples. We are able to cover about 70% of the material in this book (excluding Chapters 2 and 15). The optional material is indicated by asterisks which we typically do not cover. Both these courses require student team projects. The project in the first course usually requires a two-group comparison based on a designed experiment or an observational study. The project in the second course requires developing a multiple regression model by using a large archival data base (e.g., marketing, salary, etc.). We also teach a one-quarter graduate course that covers multiple regression (Chapter 11), ANOVA (parts of Chapters 12 and 13), and selected mathematical statistics topics from Chapter 15.

Below we give some suggested syllabi for different undergraduate courses. These assume that lecture hours are not used to give exams. Some instructors may choose not to omit certain indicated sections or cover them through self-reading/homeworks/problem sessions or decide to omit other sections instead. For example, instead of covering Section 4.4 on summarizing bivariate data, one may cover Section 4.5 on summarizing time-series data, or instead of covering Section 14.5 on nonparametric measures of correlation, one may cover Section 14.6 on resampling methods. Also, some instructors may want to give an elementary introduction to the topic of maximum likelihood estimation from the beginning sections of Chapter 15. In any case, we warn the instructor that the following syllabi are quite ambitious and intense in topical coverage, and not much extra material can be added to them. If anything, many of the asterisked sections may need to be omitted.

### A ONE-SEMESTER (45 ONE-HOUR LECTURES + 15 ONE-HOUR PROBLEM SESSIONS) PROBABILITY-STATISTICS COURSE

1. Chapter 1: (1 lecture)
2. Chapter 2: Sections 2.1, 2.2 (omit Section 2.2.4), 2.3, 2.4 (omit Sections 2.4.3 – 2.4.5), 2.5 (omit Sections 2.5.4, 2.5.7), 2.7 (omit Sections 2.7.3, 2.7.5, 2.7.6), 2.8 (omit Sections 2.8.3, 2.8.4), 2.9 (12 lectures)
3. Chapter 3: Sections 3.1, 3.2 (omit Section 3.2.2), 3.3 (2 lectures)
4. Chapter 4: Sections 4.1, 4.2, 4.3 (omit Section 4.3.3), 4.4 (5 lectures)

5. Chapter 5: Sections 5.1, 5.2 (3 lectures)
6. Chapter 6: Sections 6.1 (omit Section 6.1.2), 6.2, 6.3 (omit Section 6.3.4) (5 lectures)
7. Chapter 7: Sections 7.1, 7.2 (discuss Section 5.3 here) (3 lectures)
8. Chapter 8: Sections 8.1 (discuss Section 3.4 here), 8.2, 8.3 (4 lectures)
9. Chapter 9: Sections 9.1 (omit Section 9.1.3), 9.2 (omit Section 9.2.2) (2 lectures)
10. Chapter 10: Sections 10.1, 10.2, 10.3 (discuss Section 5.4 here), 10.4 (6 lectures)
11. Chapter 11: Sections 11.1, 11.2 (2 lectures)

**A ONE-QUARTER (30 ONE-HOUR LECTURES + 10 ONE-HOUR  
PROBLEM SESSIONS) FIRST STATISTICS COURSE WITH A PROBABILITY  
PREREQUISITE**

1. Chapter 1: (1 lecture)
2. Chapter 3: Sections 3.1, 3.2 (omit Section 3.2.2), 3.3 (2 lectures)
3. Chapter 4: Sections 4.1, 4.2, 4.3 (omit Section 4.3.3), 4.4 (5 lectures)
4. Chapter 5: Sections 5.1, 5.2 (3 lectures)
5. Chapter 6: Sections 6.1 (omit Section 6.1.2), 6.2, 6.3 (omit Section 6.3.4) (5 lectures)
6. Chapter 7: Sections 7.1, 7.2 (discuss Section 5.3 here) (3 lectures)
7. Chapter 8: Sections 8.1 (discuss Section 3.4 here), 8.2, 8.3 (4 lectures)
8. Chapter 9: Sections 9.1 (omit Section 9.1.3), 9.2 (omit Section 9.2.2) (2 lectures)
9. Chapter 10: Sections 10.1, 10.2, 10.3 (discuss Section 5.4 here) (5 lectures)

**A ONE-QUARTER (30 ONE-HOUR LECTURES + 10 ONE-HOUR  
PROBLEM SESSIONS) SECOND STATISTICS COURSE**

1. Chapter 11: Sections 11.1, 11.2, 11.3, 11.4, 11.5 (discuss Section 10.4 here), 11.6 (omit Sections 11.6.4, 11.6.5), 11.7, 11.8 (10 lectures)
2. Chapter 12: Sections 12.1, 12.2 (omit Section 12.2.2), 12.4 (omit Section 12.4.5) (6 lectures)
3. Chapter 13: Sections 13.1 (omit Sections 13.1.5, 13.1.6), 13.2 (omit Section 13.2.5) (5 lectures)
4. Chapter 9: Sections 9.3, 9.4 (omit Section 9.4.3) (3 lectures)
5. Chapter 14: Sections 14.1, 14.2 (omit Section 14.2.2), 14.3 (omit Section 14.3.2), 14.4 (omit Section 14.4.2), 14.5 (omit Section 14.5.3) (6 lectures)

**ACKNOWLEDGMENTS**

First we would like to acknowledge the source of inspiration for this book, which are the following three outstanding texts: *Statistics* by David Freedman, Robert Pisani, Roger Purves and Ani Adhikari, *Introduction to the Practice of Statistics* by David Moore and George McCabe, and *Mathematical Statistics and Data Analysis* by John Rice. Their presentation of the subject is very much in line with our view of how statistics ought to be taught (albeit at increasingly higher mathematical levels). Unfortunately, these texts did not always suit our needs because of the special backgrounds of our students and the topics that we wanted to cover in our courses. We believe we are not unique

in this respect. In trying to emulate these texts, we realize that we have set ourselves a very difficult but laudable goal.

We have taken data for exercises and presentation techniques from many other books. It is impossible to list them all, although we have tried to give credit in specific cases. We apologize for any inadvertent omissions. One group whose debt can never be fully repaid is our teachers from whom we have learned a great deal, and whose ideas and thoughts permeate this book.

We thank the following reviewers for their comments and suggestions, which have greatly improved the book: David Allen (University of Kentucky), George Casella (Cornell University), Mahtash Esfandiari (UCLA), Dale Everson (University of Idaho), Chris Franklin (University of Georgia), Robert Gould (UCLA), Steven Jordan (University of Illinois at Chicago), Carl Spruill (Georgia Tech), Gunther Walther (Stanford), and Michael Wolf (UCLA). Joe Feinglass and Yi-Cheng Hsieh of the Northwestern University Medical School provided comments on Chapters 3 and 6, respectively. Yi-Cheng Hsieh also helped in writing the solutions for the exercises from Chapter 6.

Brent Logan, a PhD student in the Department of Statistics at Northwestern University, wrote the Instructor's Solution Manual and the answers to selected odd-numbered exercises given in Appendix B and found many errors in the process. All solutions were checked by Priscilla Gathoni. Several batches of students at Northwestern University toiled through early drafts of the book and gave us invaluable feedback. An early draft was also used in a course at UCLA; comments from the instructor and students in that course were very helpful. Despite all the comments and corrections that we received from many individuals, there are bound to be some residual errors for which we take full responsibility.

We thank Professor Bruce Ankenman of the Department of Industrial Engineering and Management Sciences at Northwestern University, who helped in creating the illustration shown on the cover. We also thank Mark Harms, a research assistant in the Institute for Health Services Research and Policy Studies at Northwestern University, who converted all figures into encapsulated postscript files.

It was a pleasure to work with Ann Heath, Editor, and Betsy Williams, Production Editor, both of Prentice Hall. Ann helped us in the development of the manuscript over several years with patience and encouragement. She also converted all data files into MINITAB and SPSS formats. Betsy efficiently oversaw the transition of the manuscript into the final book form. Last but not the least, we are grateful for the love and support of our respective families. This book is dedicated to them.

Ajit C. Tamhane, Evanston, IL  
Dorothy D. Dunlop, Evanston, IL