COMPREHENSIVE VERSION

THE PRACTICE OF BUSINESS STATISTICS

USING DATA FOR DECISIONS

Includes All 18 Chapters

See back cover or Preface for other options

DAVID S. MOORE

GEORGE P. McCABE

WILLIAM M. DUCKWORTH

STANLEY L. SCLOVE

THE PRACTICE OF BUSINESS STATISTICS

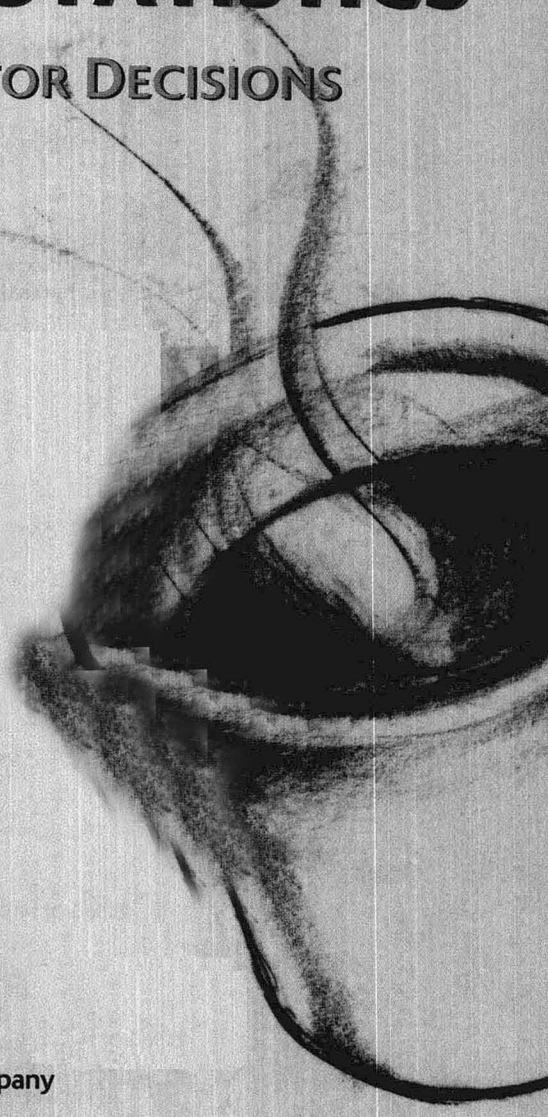
USING DATA FOR DECISIONS

David S. Moore
Purdue University

George P. McCabe
Purdue University

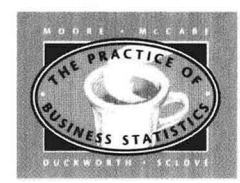
William M. Duckworth
Iowa State University

Stanley L. Sclove
University of Illinois
at Chicago



W. H. Freeman and Company
New York

Senior Acquisitions Editor :	**Patrick Farace**
Senior Developmental Editor :	**Terri Ward**
Associate Editor :	**Danielle Swearengin**
Media Editor :	**Brian Donnellan**
Marketing Manager :	**Jeffrey Rucker**
Head of Strategic Market Development :	**Clancy Marshall**
Executive Editor :	**Craig Bleyer**
Project Editor :	**Mary Louise Byrd**
Cover and Text Design :	**Vicki Tomaselli**
Cover and Interior Illustrations :	**Janet Hamlin**
Production Coordinator :	**Paul W. Rohloff**
Composition :	**Publication Services**
Manufacturing :	**RR Donnelley & Sons Company**



Library of Congress Control Number: 2002108463

ISBN-13: 978-0-7167-5723-8
ISBN-10: 0-7167-5723-0

Printed in the United States of America

Third Printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke
RG21 6XS, England
www.whfreeman.com

TO INSTRUCTORS: ABOUT THIS BOOK

Statistics is the science of data. ***The Practice of Business Statistics (PBS)*** is an introduction to statistics for students of business and economics that is based on that simple principle. Business statistics texts have tended to emphasize probability and inference. ***PBS*** reflects the current consensus among statisticians, in which data analysis and the design of data production join probability-based inference as major content areas.[1]* As the joint curriculum committee of the American Statistical Association and the Mathematical Association of America said, "Almost any course in statistics can be improved by more emphasis on data and concepts, at the expense of less theory and fewer recipes."

There are good reasons for giving more attention to data analysis and data production, along with a full treatment of inference. Data relevant to business decisions often represent complete populations for which data analysis rather than inference from sample to population is appropriate. Demographic data from the census for comparing several potential retail outlet locations have this nature, and the personnel files of a company include records for all employees. The currently hot topic of "data mining" combines the data analysis way of thinking with algorithmic advances that make exploration of immense data sets feasible. Moreover, because real data are often messy and inference requires clean data, data analysis is an essential preliminary to inference.

Themes of This Book

Look at your data is a consistent theme in ***PBS.*** Rushing to inference—often automated by software—without first exploring the data is the most common source of statistical errors that we see in working with users from many fields. A second theme is that *where the data come from matters.* When we do statistical inference, we are acting as if the data come from properly randomized sample or experimental designs. A basic understanding of these designs helps students grasp how inference works. The distinction between observational and experimental data helps students understand the truth of the mantra that "association does not imply causation." Moreover, managers need to understand the use of sample surveys for market research and customer satisfaction and of statistically designed experiments for product development, as in clinical trials of pharmaceuticals. Another strand that runs through ***PBS*** is that *data lead to decisions* in a specific setting. A calculation or graph or "reject H_0" is not the conclusion of an exercise in statistics. We encourage students to state a conclusion in the specific problem context, even though quite simple, and we hope that you will require them to do so. Finally, we think that a first course in any discipline should *focus on the essentials.* We have not tried to write an encyclopedia, but to equip students to use statistics (and learn more statistics as needed) by presenting

*All notes are collected in the Notes and Data Sources section at the end of the book.

the major concepts and most-used tools of the discipline. Longer lists of procedures "covered" tend to reduce student understanding and ability to use any procedures to deal with real problems.

Content and Style

PBS adapts to the business statistics setting the approach to introductory instruction that was inaugurated and proved successful in the best-selling general statistics texts *Introduction to the Practice of Statistics* (fourth edition, Freeman 2003) and *The Basic Practice of Statistics* (third edition, Freeman, 2004). Like these books, ***PBS*** features use of real data in examples and exercises and emphasizes statistical thinking as well as mastery of techniques. As the continuing revolution in computing automates most tiresome details, an emphasis on statistical concepts and on insight from data becomes both more practical for students and teachers and more important for users who must supply what is not automated.

Chapters 1 and 2 present the methods and unifying ideas of *data analysis.* Students appreciate the usefulness of data analysis, and that they can actually do it relieves a bit of their anxiety about statistics. We hope that they will grow accustomed to examining data and will continue to do so even when formal inference to answer a specific question is the ultimate goal. Note in particular that Chapter 2 gives an extended treatment of *correlation and regression* as descriptive tools, with attention to issues such as influential observations and the dangers posed by lurking variables, These ideas and tools have wider scope than an emphasis on inference (Chapters 10 and 11) allows. We think that a full discussion of data analysis for both one and several variables before students meet inference in these settings both reflects statistical practice and is pedagogically helpful.

Teachers will notice some nonstandard ideas in these chapters, particularly regarding the *Normal distributions*—we capitalize "Normal" to avoid suggesting that these distributions are "normal" in the usual sense of the word. We introduce density curves and Normal distributions in Chapter 1 as models for the overall pattern of some sets of data. Only later (Chapter 4) do we see that the same tools can describe probability distributions. Although unusual, this presentation reflects the historical origin of Normal distributions and also helps break up the mass of probability that is so often a barrier students fail to surmount. We use the notation $N(\mu, \sigma)$ rather than $N(\mu, \sigma^2)$ for Normal distributions. The traditional notation is in fact indefensible other than as inherited tradition The standard deviation, not the variance, is the natural measure of scale in Normal distributions, visible on the density curve, used in standardization, and so on. We want students to think in terms of mean and standard deviation, so we talk in these terms.

In Chapter 3, we discuss *random sampling* and *randomized comparative experiments*. The exposition pays attention to practical difficulties, such as nonresponse in sample surveys, that can greatly reduce the value of data. We think that an understanding of such broader issues is particularly important for managers who must use data but do not themselves produce data. Discussion of statistics in practice alongside more technical material is part of our emphasis on data leading to practical decisions.

Chapter 3 also uses the idea of random sampling to motivate the need for statistical inference (sample results vary) and probability (patterns of random variation) as the foundation for inference. Chapters 4 and 5 then present *probability*. We have chosen an unusual approach: Chapter 4 contains only the probability material that is needed to understand statistical inference, and this material is presented quite informally. Chapter 5 presents additional probability in a more traditional manner. *Chapter 4 is required to read the rest of the book, but Chapter 5 is optional.* We suggest that you consider omitting Chapter 5 unless your students are well prepared or have some need to know probability beyond an understanding of basic statistics. One reason is to maintain content balance—less time spent on formal probability allows full attention to data analysis without reducing coverage of inference. Pedagogical concerns are more compelling. Experienced teachers recognize that students find probability difficult. Research on learning confirms our experience. Even students who can do formally posed probability problems often have a very fragile conceptual grasp of probability ideas.[2] Formal probability does not help students master the ideas of inference (at least not as much as we teachers imagine), and it depletes reserves of mental energy that might better be applied to essentially statistical ideas.

The remaining chapters present *statistical inference,* still encouraging students to ask where the data come from and to look at the data rather than quickly choosing a statistical test from an Excel menu. Chapter 6, which describes the *reasoning of inference,* is the cornerstone. Chapters 7 and 8 discuss *one-sample and two-sample procedures,* which almost any first course will cover. We take the opportunity in these core "statistical practice" chapters to discuss practical aspects of inference in the context of specific examples. Chapters 9, 10, and 11 present selected more advanced topics in inference: *two-way tables* and *simple and multiple regression.*

Instructors who wish to customize a single-semester course or to add a second semester will find a wide choice of additional topics in the economical paperbound *Companion Chapters* that extend ***PBS.*** These chapters are:

Chapter 12 Statistics for Quality: Control and Capability

Chapter 13 Time Series Forecasting

Chapter 14 One-Way Analysis of Variance

Chapter 15 Two-Way Analysis of Variance

Chapter 16 Nonparametric Tests

Chapter 17 Logistic Regression

Chapter 18 Bootstrap Methods and Permutation Tests

Companion Chapters can be ordered individually or packaged in flexible combinations with the Core book. A Comprehensive Version of all 18 chapters is also available.

One last comment on the style and content of ***PBS.*** We hope that the book presents an up-to-date picture of statistical practice, at least as far as such a picture is within the reach of beginners. Thus, for example, we encourage

use of the version of the two-sample t procedures that does not assume equal population variances. We present the modified ("add two successes and two failures") confidence intervals for proportions that are now supported both by extensive simulation and by theory. We point out that p charts for process control are unsuitable for the high quality levels typical of modern manufacturing, but remain useful for many business processes. We include brief optional "Beyond the Basics" overviews of newer methods, such as density estimation. scatterplot smoothers, capture-recapture sampling, data mining, and the bootstrap.

Accessible Technology

Any mention of the current state of statistical practice reminds us that quick, cheap, and easy computation has changed the field. Procedures such as our recommended two-sample t and logistic regression, not to mention the bootstrap and data mining, depend on software. Even the mantra "look at your data" depends in practice on software, as making multiple plots by hand is too tedious when quick decisions are required. What is more, automating calculations and graphs increases students' ability to complete problems, reduces their frustration, and helps them concentrate on ideas and problem recognition rather than mechanics.

*We therefore strongly recommend that a course based on **PBS** be accompanied by software of your choice.* Instructors will find using software easier because all data sets for ***PBS*** can be found in several common formats both on the Web (`www.whfreeman.com/pbs`) and on the CD-ROM that accompanies each copy of the book.

The Microsoft Excel spreadsheet is by far the most common program used for statistical analysis in business. Our displays of output therefore emphasize Excel, though output from several other programs also appears. ***PBS*** is not tied to specific software and does not give instruction in using software. (Separate manuals linked to ***PBS*** are available to guide the learning of several common software systems—see the description of supplements on pages xxvi and xxviii.) Indeed, one of our emphases is that a student who has mastered the basics of, say, regression can interpret and use regression output from almost any software. Figure 2.13 (page 115) displays regression output from Excel, the Minitab statistical software, and the TI-83 graphing calculator to illustrate this point. Similar displays appear elsewhere in the book.

We are well aware that Excel lacks many advanced statistical procedures. More seriously, Excel's statistical procedures have been found to be inaccurate, and they lack adequate warnings for users when they encounter data for which they may give incorrect answers.[3] There is good reason for people whose profession requires continual use of statistical analysis to avoid Excel. But there are also good practical reasons why managers whose life is not statistical prefer a program that they regularly use for other purposes. Excel appears to be adequate for simpler analyses of the kind that occur most often in business applications.

Some statistical work, both in practice and in ***PBS,*** can be done with a calculator rather than software. *Students should have at least a "two-*

variable statistics" calculator with functions for correlation and the least-squares regression line as well as for the mean and standard deviation. Graphing calculators offer considerably more capability. Because students have calculators, the text doesn't discuss "computing formulas" for the sample standard deviation or the least-squares regression line.

Technology can be used to assist *learning* statistics as well as *doing* statistics. The design of good software for learning is often quite different from that of software for doing. We want to call particular attention to the set of statistical applets available on the ***PBS*** Web site: `www.whfreeman.com/pbs`. These interactive graphical programs are by far the most effective way to help students grasp the sensitivity of correlation and regression to outliers, the idea of a confidence interval, the way ANOVA responds to both within-group and among-group variation, and many other statistical fundamentals. Exercises using these applets appear throughout the text, marked by a distinctive icon. We urge you to assign some of these, and we suggest that if your classroom is suitably equipped, the applets are very helpful tools for classroom presentation as well.

Carefully Structured Pedagogy

Few students find statistics easy. An emphasis on real data and real problems helps maintain motivation, and there is no substitute for clear writing. Beginning with data analysis builds confidence and gives students a chance to become familiar with your chosen software before the statistical content becomes intimidating. We have adopted several structural devices to aid students. Major settings that drive the exposition are presented as *cases* with more background information than other examples. (But we avoid the temptation to give so much information that the case obscures the statistics.) A distinctive icon ties together examples and exercises based on a case.

The *exercises* are structured with particular care. Short "Apply Your Knowledge" sections pose straightforward problems immediately after each major new idea. These give students stopping points (in itself a great help to beginners) and also tell them that "you should be able to do these things right now." Each numbered section in the text ends with a substantial set of exercises, and more appear as review exercises at the end of each chapter. Finally, each chapter ends with a few "Case Study Exercises" that are suitable for individual or group projects. Case Study Exercises are more ambitious, offer less explicit guidance, and often use large data sets.

Acknowledgments

We are grateful to the many colleagues and students who have provided helpful comments about *Introduction to the Practice of Statistics* and *The Basic Practice of Statistics*. They have contributed to improving ***PBS*** as well. In particular, we would like to thank the following colleagues who, as reviewers and class testers, offered specific comments on ***PBS:***

Mohamed H. Albohali,
Indiana University of Pennsylvania
Mary Alguire,
University of Arkansas
Andrew T. Allen,
University of San Diego
Djeto Assane,
University of Nevada–Las Vegas
Lynda L. Ballou,
Kansas State University
Ronald Barnes,
University of Houston–Downtown
Paul Baum,
California State University–Northridge
Vanessa Beddo,
University of California–Los Angeles
Dan Brick,
University of St. Thomas, St. Paul–Minnesota
Alan S. Chesen,
Wright State University
Siddhartha Chib,
Washington University
Judith Clarke,
California State University–Stanislaus
Lewis Coopersmith,
Rider University
Jose Luis Guerrero Cusumano,
Georgetown University
Frederick W. Derrick,
Loyola College, Maryland
Zvi Drezner,
California State University–Fullerton
Bill Duckworth,
Iowa State University
Abdul Fazal,
California State University–Stanislaus
Yue Feng,
University of Oregon
Paul W. Guy,
California State University–Chico
Robert Hannum,
University of Denver
Erin M. Hodgess,
University of Houston–Downtown
J. D. Jobson,
University of Alberta
Howard S. Kaplon,
Towson University
Mohyeddin Kassar,
University of Illinois–Chicago
Michael L. Kazlow,
Pace University
Nathan R. Keith,
Devry University
John F. Kottas,
College of William and Mary
Linda S. Leighton,
Fordham University
Ramon V. Leon,
University of Tennessee–Knoxville
Ben Lev,
University of Michigan–Dearborn
Vivian Lew,
University of California–Los Angeles
Gene Lindsay,
St. Charles Community College
Richard N. Madsen,
University of Utah
Roberto S. Mariano,
University of Pennsylvania
David Mathiason,
Rochester Institute of Technology
B. D. McCullough,
Drexel University
John D. McKenzie, Jr.,
Babson College
Tim Novotny,
Mesa State College
Tom Obremski,
University of Denver
J. B. Orris,
Butler University
Steve Ramsier,
Florida State University
Stephen Reid,
BC Institute of Technology
Ralph Russo,
University of Iowa
Neil C. Schwertman,
California State University–Chico
Carlton Scott,
University of California–Irvine
Thomas R. Sexton,
SUNY–Stony Brook
Tayyeb Shabbir,
University of Pennsylvania

Anthony B. Sindone,
University of Notre Dame
John Sparks,
University of Illinois–Chicago
Michael Speed,
Texas A&M University
Debra Stiver,
University of Nevada–Reno
Sandra Strasser,
Valparaiso University
David Thiel,
University of Nevada–Las Vegas
Milan Velebit,
University of Illinois–Chicago
Raja Velu,
Syracuse University
Robert J. Vokurka,
Texas A&M University–Corpus Christi
Art Warburton,
Simon Fraser University
Jay Weber,
University of Illinois–Chicago
Rodney M. Wong,
University of California–Davis
Elaine Zanutto,
University of Pennsylvania
Zhe George Zhang,
Western Washington University

The professionals at W. H. Freeman and Company, in particular Mary Louise Byrd, Patrick Farace, and Terri Ward, have contributed greatly to the success of ***PBS.*** Most of all, we are grateful to the many people in varied disciplines and occupations with whom we have worked to gain understanding from data. They have provided both material for this book and the experience that enabled us to write it. What the eminent statistician John Tukey called "the real problems experience and the real data experience" has shaped our view of statistics. It has convinced us of the need for beginning instruction to focus on data and concepts, building intellectual skills that transfer to more elaborate settings and remain essential when all details are automated. We hope that users and potential users of statistical techniques will find this emphasis helpful.

David S. Moore
George P. McCabe
William M. Duckworth
Stanley L. Sclove

SUPPLEMENTS AND MEDIA FOR STUDENTS

A full range of supplements and media is available to help students get the most out of *PBS:*

Supplements

- **Student Solutions Manual** (0-7167-9860-3), prepared by Michael A. Fligner and William I. Notz of The Ohio State University, offers students explanations of crucial concepts in each section of ***PBS***, plus detailed solutions to the odd-numbered exercises and step-through models of important statistical techniques.
- **Statistical Software Manuals** will guide students in the use of particular statistical software with ***PBS***. The chapters of each manual correlate with those of ***PBS*** and include a set of exercises specific to that chapter's concepts. These manuals are:
  - **Excel Manual** (0-7167-6640-X), developed by Fred Hoppe of McMaster University. Excel Macros can be accessed on our Web site, `www.whfreeman.com/pbs`
  - **JMP Manual** (0-7167-9630-9), prepared by Thomas Devlin of Montclair State University.
  - **Minitab Manual** (0-7167-9787-9), prepared by Betsy Greenberg and Pallavi Chitturi of the University of Texas at Austin.
  - **SPSS Manual** (0-7167-9690-2), prepared by James A. Danowski of the University of Illinois at Chicago.
- **TI-83 Graphing Calculator Manual** (0-7167-9691-0), prepared by David K. Neal of Western Kentucky University.
- **Case Book** (0-7167-5747-8), prepared by William I. Notz, Dennis K. Pearl, and Elizabeth Stasny of The Ohio State University, offers a variety of additional, in-depth case studies that can be utilized as homework or group activities. All case studies are business-related and contain real data.
- **Projects Book** (0-7167-9809-3), prepared by Ron Millard of Shawnee Mission South High School and John C. Turner of the U.S. Naval Academy, offers students, as well as instructors, ideas for hands-on explorations that prompt students to think critically about statistics.

Media

- ***PBS* Web site** `www.whfreeman.com/pbs` seamlessly integrates topics from the text. On the Web site students can find:
  - **Interactive Statistical Applets** that allow students to manipulate data and see results graphically. These applets automate calculations and graphics in a way that nicely complements classroom work.

- **Data Sets** in ASCII, JMP, Minitab, SPSS, TI and Excel formats.
- **Self Quizzes** to help students prepare for tests. Each quiz contains 20 multiple-choice questions that include automatic feedback for incorrect answers, reinforcement for correct answers, and text section references to aid in additional study for incorrect answers.
- **Student version of the Electronic Encyclopedia of Statistical Examples and Exercises (EESEE)** is a rich repository of case studies that apply the concepts of the text to various real-world venues, such as the mass media, business, sports, natural sciences, social sciences, and medicine. Each case study is accompanied by practice problems, and most include full data sets that are exportable to various statistical software packages. EESEE was developed by a consortium at The Ohio State University dedicated to statistical education.

- **Excel Macros,** developed by Fred Hoppe of McMaster University. These macros facilitate use of spreadsheet operations by students and instructors, allowing for easier and more accurate data and statistical analyses. Adopters of the **Excel Manual** will be able to download updates to these macros.
- **Statistics** have been gathered from key business sources on the World Wide Web and placed in **Excel spreadsheets.** Exercises based on these spreadsheets have been written that test students' understanding of business data in Excel. These exercises maintain currency by being updated every three months.
- **Interactive Student CD-ROM** is included with every new copy of ***PBS.*** Note that the CD covers material for Chapters 1–11. Materials for Chapters 12–18 can be found on the book's Web site: `www.whfreeman.com/pbs`. The data sets, however, for all 18 chapters are contained on the CD. The CD also contains:
  - **EESEE (Electronic Encyclopedia of Statistical Examples and Exercises)** case studies.
  - **Self Quizzes** in multiple-choice format for each chapter constructed by experienced instructors to anticipate typical errors.
  - **Data Sets** in ASCII, JMP, Minitab, TI, SPSS, and Excel formats.
  - **Interactive Statistical Applets** that allow students to manipulate data and see results graphically. These applets automate calculations and graphics in a way that nicely complements classroom work.
  - **Student versions of Minitab, SPSS, JMP, and SPLUS** can be packaged with ***PBS.***
  - **Brief Minitab and Excel instructions** on how these softwares can be used to further explore and understand each chapter's content. The Excel Instructions were prepared by Fred Hoppe of McMaster University, and the Minitab Instructions were prepared by Betsy Greenberg and Pallavi Chitturi of the University of Texas at Austin.

SUPPLEMENTS AND MEDIA FOR INSTRUCTORS

A full range of supplements and media support is available to help instructors better teach from ***PBS:***

Instructor's CD

Instructor's Resource CD-ROM (0-7167-9842-5) contains all the student CD material plus the following:

- **Instructor's Guide with Solutions** in Adobe .pdf electronic format.
- **Instructor's version of the Electronic Encyclopedia of Statistical Examples and Exercises (EESEE),** with solutions to the exercises in the student version.
- **Self Quizzes** in multiple-choice format for each chapter constructed by experienced instructors to anticipate typical errors.
- **Data Sets** in ASCII, JMP, Minitab, TI, SPSS, and Excel formats.
- **Interactive Statistical Applets** that allow students to manipulate data and see results graphically. These applets automate calculations and graphics in a way that nicely complements classroom work.
- **All *PBS* figures** in an exportable presentation format, JPEG for Windows and PICT for Macintosh users.
- **Presentation Manager Pro™,** which creates presentations for all the figures and selected tables from the text. No extra software is necessary to save, print, and present electronic presentations using the materials from the text. Extra material can be imported to the presentation list from locally saved files and the Web.
- **PowerPoint™ Slides** that can be used directly or customized for your particular course. Every image and table from the textbook is formatted for your lecture needs.
- **Brief Minitab and Excel instructions** on how these softwares can be used to further explore and understand each chapter's content. The Excel Instructions were prepared by Fred Hoppe of McMaster University, and the Minitab Instructions were prepared by Betsy Greenberg and Pallavi Chitturi of the University of Texas at Austin.

Assessment Tools

- **Instructor's Guide with Solutions** (0-7167-9692-9), prepared by Lori Seward of the University of Colorado at Boulder, includes worked-out solutions to all exercises, teaching suggestions, and chapter comments.
- **Test Bank** (0-7167-6641-8), prepared by Michael Fligner and William Notz of The Ohio State University, is an easy-to-use CD that includes Windows and Mac versions on a single disc. The format lets you add,

edit, and resequence questions to suit your needs. This is also available as a print supplement (0-7167-6642-6).

- **Online Testing** powered by Diploma from the Brownstone Research Group offers instructors the ability to easily create and administer secure exams over a network and over the Internet, with questions that incorporate multimedia and interactive exercises. The program lets you restrict tests to specific computers or time blocks, and includes an impressive suite of gradebook and result-analysis features.
- **Online Quizzing** powered by Question Mark and accessed via the ***PBS*** Web site uses Question Mark's Perception to enable instructors to easily and securely quiz students online using prewritten, multiple-choice questions for each text chapter, separate from those appearing in the *Test Bank*. Students receive instant feedback and can take the quizzes multiple times. Instructors can view the results by quiz, student, or question or can get weekly results via e-mail.

PBS Web Site

The ***PBS* Web site for Instructors,** `www.whfreeman.com/pbs` contains all features available to students, plus the following:

- **Instructor's version of the Electronic Encyclopedia of Statistical Examples and Exercises (EESEE),** with solutions to the exercises in the student version.
- **PowerPoint Slides** that can be used directly or customized for your particular course. Every image and table from the textbook is formatted for your lecture needs.

Course Management

- **Online Course Materials (WebCT and Blackboard)** can be provided as a service for adopters. We offer electronic content of ***PBS,*** including the complete *Test Bank* and all Web site content in WebCT and Blackboard.

Statistical Software Packages

- **Student versions of Minitab, SPSS, JMP, and SPLUS** can be bundled with ***PBS*** for those instructors who wish to use a statistical software package in the course.

TO STUDENTS: WHAT IS STATISTICS?

Statistics is the science of collecting, organizing, and interpreting numerical facts, which we call *data.*

We are bombarded by data in our everyday life. The news mentions imported car sales, the latest poll of the president's popularity, and the average high temperature for today's date. Advertisements claim that data show the superiority of the advertiser's product. All sides in public debates about economics, education, and social policy argue from data. A knowledge of statistics helps separate sense from nonsense in the flood of data.

An understanding of data is also a key skill for business managers. Market research that reveals consumer tastes guides development of products and services. Data from retail outlets, if looked at with understanding, can suggest better ways to position products in diverse markets. The success of new product lines or new advertising campaigns is assessed from data. Manufacturers improve their products and processes by using data on their quality and reliability. Monthly government data on unemployment, inflation, and economic growth influence strategic business decisions. The study of investment portfolios has become so heavily statistical that standard deviations and correlations appear even in mutual fund reports. If your company's personnel data appear to show that women or minorities are not fairly treated, you can expect a lawsuit that will require very thorough study of employee data. And don't forget that firms from General Motors to some of the world's largest banks have suffered losses of millions of dollars from fraud or unauthorized transactions that would have been obvious if managers (now former managers) had looked at the data presented to them.

Understanding from Data

The goal of statistics is to gain understanding from data. To gain understanding, we often operate on a set of numbers—we average or graph them, for example. But we must do more, because data are not just numbers; they are *numbers with a context* that helps us understand them. You hear that this month's unemployment rate is 6%. What does this number mean? Will it pose political problems for the party in power? Is it likely to reduce consumer spending and so reduce sales of your firm's products or services?

Much of the context rests on comparing this number with past data. An unemployment rate of 6% is somewhat high for the United States, where unemployment averaged 5.4% in the decade from 1993 to 2002. But 6% is low for Spain, where unemployment rates during the same decade ranged from over 10% to more than 22%. The president must worry about 6% unemployment in the United States, but a Spanish prime minister could win reelection for so low a rate.

The way variables are measured is also part of the context. A nation's unemployment rate is not a percent of all working-age people, but of the labor force—people who are available for work and looking for work. If public policy shrinks the labor force by offering early retirement pensions or

making disability payments easily available, the unemployment rate will go down. Some European nations have done this. The context of a number also includes the methods used to produce it. Unemployment data in developed nations come from national sample surveys of households. That is, many households chosen by chance are asked about the employment status of all adults living there. This is superior to the older method of counting only people who signed up for unemployment benefits, which missed all those who were not eligible or didn't trouble to sign up.

When you do statistical problems—even simple textbook problems—don't just graph or calculate. Think about the context and state your conclusions in the specific setting of the problem. As you are learning how to do statistical calculations and graphs, remember that the goal of statistics is not calculation for its own sake but gaining understanding from numbers. The calculations and graphs can be automated by a calculator or software, but you must supply the understanding. This book presents only the most common specific procedures for statistical analysis. A thorough grasp of the principles of statistics will enable you to quickly learn more advanced methods as needed. On the other hand, a fancy computer analysis carried out without attention to basic principles will often produce elaborate nonsense. As you read, seek to understand the principles as well as the necessary details of methods and recipes.

The Rise of Statistics

Historically, the ideas and methods of statistics developed gradually as society grew interested in collecting and using data for a variety of applications. The earliest origins of statistics lie in the desire of rulers to count the number of inhabitants or measure the value of taxable land in their domains. As the physical sciences developed in the seventeenth and eighteenth centuries, the importance of careful measurements of weights, distances, and other physical quantities grew. Astronomers and surveyors striving for exactness had to deal with variation in their measurements. Many measurements should be better than a single measurement, even though they vary among themselves. How can we best combine many varying observations? Statistical methods that are still important were invented in order to analyze scientific measurements.

By the nineteenth century, the agricultural, life, and behavioral sciences also began to rely on data to answer fundamental questions. How are the heights of parents and children related? Does a new variety of wheat produce higher yields than the old, and under what conditions of rainfall and fertilizer? Can a person's mental ability and behavior be measured just as we measure height and reaction time? Effective methods for dealing with such questions developed slowly and with much debate.[1]

In the twentieth century, economics, finance, and the analysis of business decisions became heavily quantitative. Ideas and techniques that originated in the collection of government data, in the study of astronomical or biological measurements, and in the attempt to understand heredity or intelligence were used to describe national economies and to study investment portfolios.

By this time, these ideas and techniques had come together to form a unified "science of data." That science of data—statistics—is the topic of this book.

The Organization of This Book

Part I of this book, called simply "Data," concerns data analysis and data production. The first two chapters deal with statistical methods for organizing and describing data. These chapters progress from simpler to more complex data. Chapter 1 examines data on a single variable; Chapter 2 is devoted to relationships among two or more variables. You will learn both how to examine data produced by others and how to organize and summarize your own data. These summaries will be first graphical, then numerical, then, when appropriate, in the form of a mathematical model that gives a compact description of the overall pattern of the data. Chapter 3 outlines arrangements (called designs) for producing data that answer specific questions. The principles presented in this chapter will help you to design proper samples and experiments and to evaluate such investigations when you are presented with their results.

Part II, consisting of Chapters 4 to 8, introduces statistical inference—formal methods for drawing conclusions from properly produced data. Statistical inference uses the language of probability to describe how reliable its conclusions are, so some basic facts about probability are needed to understand inference. Probability is the subject of Chapter 4 and the optional Chapter 5. Chapter 6, perhaps the most important chapter in the text, introduces the reasoning of statistical inference. Effective inference is based on good procedures for producing data (Chapter 3), careful examination of the data (Chapters 1 and 2), and an understanding of the nature of statistical inference as discussed in Chapter 6. Chapters 7 and 8 describe some of the most common specific methods of inference, for drawing conclusions about means and proportions from one and two samples.

The three shorter chapters in Part III introduce somewhat more advanced methods of inference, dealing with relations in categorical data and regression and correlation. Your instructor may also choose to have you study one or more of the separately bound Companion Chapters that present additional statistical topics.

What Lies Ahead

The Practice of Business Statistics is full of data. Many exercises ask you to express briefly some understanding gained from the data. In practice, you would know much more about the background of the data you work with and about the questions you hope the data will answer. No textbook can be fully realistic. But it is important to form the habit of asking, "What do the data tell me?" rather than just concentrating on making graphs and doing calculations.

You should have some help in automating many of the graphs and calculations. You should certainly have a calculator with basic statistical