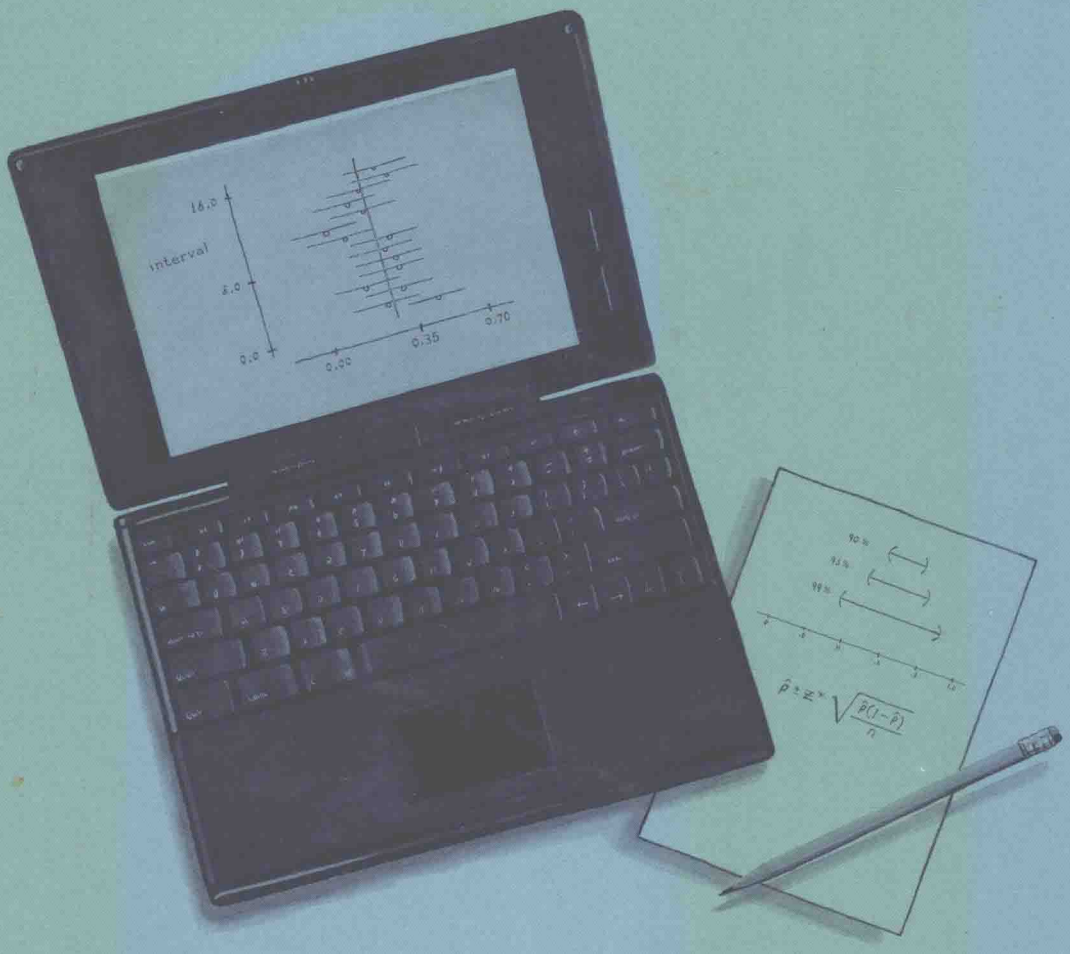
THE WORKSHOP MATHEMATICS PROJECT

WORKSHOP STATISTICS

DISCOVERY WITH DATA AND MINITAB®



ALLAN J. ROSSMAN BETH L. CHANCE

Workshop Statistics

Discovery with Data and Minitab®

Allan J. Rossman

DICKINSON COLLEGE

Beth L. Chance

UNIVERSITY OF THE PACIFIC

Springer

Library of Congress Cataloging-in-Publication Data
Rossman, Allan J.
Workshop statistics : discovery with data and Minitab/Alan J. Rossman, Beth Chance.
p. cm. — (Textbooks in mathematical sciences)
Includes bibliographical references (p.)
ISBN 0-387-98411-9 (softcover : alk. paper)
1. Mathematical Statistics. 2. Minitab. I. Chance, Beth L. II. Title. III. Series.
QA276.12.R6734 1998
519.5—dc21 97-52162

Printed on acid-free paper.

Production managed by Lesley Poliner; manufacturing supervised by Joe Quatela.
Photocomposed by Matrix Publishing Services, York, PA.
Printed and bound by Hamilton Printing Co., Rensselaer, NY.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-98411-9 Springer-Verlag New York Berlin Heidelberg SPIN 10659932

Preface to the Minitab® Version

This version of *Workshop Statistics* incorporates instructions specific to the Minitab software (Macintosh and Windows platforms, Professional and Student versions) while maintaining all the distinctive features of the original text. The book continues to focus on active learning, conceptual understanding, and genuine data, but now efficiently integrates Minitab instructions into the activities. This union enhances students' ability to explore statistics visually, to investigate long-term behavior of sample statistics, and to focus on interpretation of results over computational manipulations.

By no means is this text intended as a user's manual for Minitab; in fact, we make use of just a fraction of Minitab's capabilities. Instead, we focus on helping students to use Minitab as a tool for analyzing data as well as an aid for discovering statistical concepts and exploring statistical principles. We chose Minitab as the software package because of its accessibility and user-friendliness to those not familiar with computers. Other appealing features of Minitab include its widespread use in both academic and industrial settings and its macro and simulation capabilities.

We intend for this text to be flexible enough to support a variety of implementations for instructors and students with varying degrees of access to computing facilities. In particular, with few exceptions we recommend that students investigate a concept through hand-drawn displays and calculator-assisted computations before proceeding to Minitab to check their work or to explore larger data sets. On most homework activities, whether or not a student should use Minitab is left to the instructor's discretion, but we provide reference to the relevant Minitab worksheets. Our goal has been to provide detailed enough instructions that students can complete Minitab assignments, with a minimum of instructor support, outside of class time if necessary. The Minitab instructions and macros have been tested with Release 11.2 for Windows, Release 9 for Windows Student, as well as Student Release 8.

While the most obvious and substantial changes from previous versions of *Workshop Statistics* involve Minitab instructions, several other updates and changes have also been made. Some of the data sets have been updated to more current data. We have also added Minitab terms to the index and a glossary where students can record important terms and definitions. Suggestions to both instructors and students for successfully implementing the workshop approach have also been included.

Support materials available for use with this text include downloadable Minitab worksheets and macros for all platforms. Our web sites (see p. xvii) also offer a variety of teaching resources ranging from an instructor's guide and answers for selected activities to sample exams and syllabi.

Acknowledgments:

I want to thank instructors around the country who have provided helpful feedback on earlier versions of *Workshop Statistics* to myself and Allan Rossman.

Jim Albert
Skip Allis
Patricia Bassett
Charles Bertness
Chuck Biehl
Jim Bohan
Cheri Boyd
Gordon Bril
Julie Clark
Benjamin Collins
Al Coons
Carolyn Cuff
Christine Czapleski
L.J. Davis
Carolyn Dobler
Clark Engel
Christa Fratto
Steve Friedberg
Brian Gray
Dorothea Grimm
Bill Halteman
Alice Hankla
Bruce King
Larry Langley
Suzanne Larson
Todd Lee
Jerry Moreno
Sue Peters
Gina Reed
Bill Rinaman
Charlie Scheim
Ned Schillow
Brian Schott
Bernie Schroeder
Joanne Schweinsberg
Sallie Scudder
Mike Seyfried
Sue Suran
Sam Tumulo
Kathryn Voit
Barr von Oehsen
Don Weimer
Rhonda Weissman
Jean Werner
Sheila Young
Thomas Zachariah
John Zhang

I am also grateful to all the students who have explored with these materials and to my own students, who have always been my motivation and inspiration.

I also thank Minitab's Author Assistance Program, particularly Christine Sarris and Jeff Hartzell, for their help in the preparation of this Minitab version.

I thank Springer-Verlag for their support, especially Jerry Lyons, Sabrina Paris, and Lesley Poliner. I also appreciate the continuing support from the Workshop Mathematics Program, directed by Nancy Baxter Hastings and managed by Joanne Weissman, and the Fund for the Improvement of Post-Secondary Education.

I would like to thank Allan Rossman for his extraordinary vision and for allowing me to share in this creative process. I also thank my husband Frank for his patience, understanding, and culinary skills. Thanks also to Tigger and Lucky for their late night company.

Beth L. Chance
University of the Pacific
January 1998

Preface

> Shorn of all subtlety and led naked out of the protective fold of educational research literature, there comes a sheepish little fact: lectures don't work nearly as well as many of us would like to think.
>
> —*George Cobb (1992)*

This book contains activities that guide students to discover statistical concepts, explore statistical principles, and apply statistical techniques. Students work toward these goals through the analysis of genuine data and through interaction with one another, with their instructor, and with technology. Providing a one-semester introduction to fundamental ideas of statistics for college and advanced high school students, *Workshop Statistics* is designed for courses that employ an interactive learning environment by replacing lectures with hands-on activities. The text contains enough expository material to stand alone, but it can also be used to supplement a more traditional textbook.

Some distinguishing features of *Workshop Statistics* are its emphases on active learning, conceptual understanding, genuine data, and the use of technology. The following sections of this preface elaborate on each of these aspects and also describe the unusual organizational structure of this text.

ACTIVE LEARNING

> Statistics teaching can be more effective if teachers determine what it is they really want students to know and to do as a result of their course, and then provide activities designed to develop the performance they desire.
>
> —*Joan Garfield (1995)*

This text is written for use with the workshop pedagogical approach, which fosters active learning by minimizing lectures and eliminating the conventional distinction between laboratory and lecture sessions. The book's ac-

tivities require students to collect data, make predictions, read about studies, analyze data, discuss findings, and write explanations. The instructor's responsibilities in this setting are to check students' progress, ask and answer questions, lead class discussions, and deliver "mini-lectures" where appropriate. The essential point is that every student is actively engaged with learning the material through reading, thinking, discussing, computing, interpreting, writing, and reflecting. In this manner students construct their own knowledge of statistical ideas as they work through the activities.

The activities also lend themselves to collaborative learning. Students can work together through the book's activities, helping each other to think through the material. Some activities specifically call for collaborative effort through the pooling of class data.

The text also stresses the importance of students' communication skills. As students work through the activities, they constantly read, write, and talk with one another. Students should be encouraged to write their explanations and conclusions in full, grammatically correct sentences, as if to an educated layperson.

CONCEPTUAL UNDERSTANDING

> Almost any statistics course can be improved by more emphasis on data and on concepts at the expense of less theory and fewer recipes. —*David Moore (1992)*

This text focuses on the "big ideas" of statistics, paying less attention to details that often divert students' attention from larger issues. Little emphasis is placed on numerical and symbolic manipulations. Rather, the activities lead students to explore the meaning of concepts such as variability, distribution, outlier, tendency, association, randomness, sampling, sampling distribution, confidence, significance, and experimental design. Students investigate these concepts by experimenting with data, often with the help of technology. Many of the activities challenge students to demonstrate their understanding of statistical issues by asking for explanations and interpretations rather than mere calculations.

To deepen students' understandings of fundamental ideas, the text presents these ideas repetitively. For example, students return to techniques of exploratory data analysis when studying properties of randomness and also in conjunction with inference procedures. They also encounter issues of data collection not just when studying randomness but also when investigating statistical inference.

GENUINE DATA

> We believe that data should be at the heart of all statistics education and that students should be introduced to statistics through data-centered courses.
>
> —*Thomas Moore and Rosemary Roberts (1989)*

The workshop approach is ideally suited to the study of statistics, the science of reasoning from data, for it forces students to be actively engaged with genuine data. Analyzing genuine data not only exposes students to the practice of statistics; it also prompts them to consider the wide applicability of statistical methods and often enhances their enjoyment of the material.

Some activities ask students to analyze data about themselves that they collect in class, while most present students with genuine data from a variety of sources. Many questions in the text ask students to make predictions about data before conducting their analyses. This practice motivates students to view data not as naked numbers but as numbers with a context, to identify personally with the data, and to take an interest in the results of their analyses.

The data sets in *Workshop Statistics* do not concentrate on one academic area but come from a variety of fields of application. These fields include law, medicine, economics, psychology, political science, and education. Many examples come not from academic disciplines but from popular culture. Specific examples therefore range from such pressing issues as testing the drug AZT and assessing evidence in sexual discrimination cases to less crucial ones of predicting basketball salaries and ranking *Star Trek* episodes.

USE OF TECHNOLOGY

> Automate calculation and graphics as much as possible.
>
> —*David Moore (1992)*

This text assumes that students have access to technology for creating visual displays, performing calculations, and conducting simulations. The preferable technology is a statistical software package, although a graphing calculator can do almost as well. Roughly half of the activities ask students to use technology. Students typically perform small-scale displays, calculations, and simulations by hand before letting the computer or calculator take over those mechanical chores.

This workshop approach employs technology in three distinct ways. First, technology performs the calculations and presents the visual displays necessary to analyze genuine data sets which are often large and cumbersome. Next, technology conducts simulations which allow students to visualize and explore the long-term behavior of sample statistics under repeated random sampling.

The most distinctive use of technology with the workshop approach is to enable students to explore statistical phenomena. Students make predictions about a particular statistical property and then use the computer to investigate their predictions, revising their predictions and iterating the process as necessary. For example, students use technology to investigate the effects of outliers on various summary statistics and the effects of sample sizes on confidence intervals.

Activities requiring the use of technology are integrated throughout the text, reinforcing the idea that technology is not to be studied for its own sake but as an indispensable tool for analyzing genuine data and a convenient device for exploring statistical phenomena.

Specific needs of the technology are to create visual displays (dotplots, histograms, boxplots, scatterplots), calculate summary statistics (mean, median, quartiles, standard deviation, correlation), conduct simulations (with binary variables), and perform inference procedures (z-tests and z-intervals for binary variables, t-tests and t-intervals for measurement variables).

ORGANIZATION

> Judge a statistics book by its exercises, and you cannot go far wrong. —*George Cobb (1987)*

For the most part this text covers traditional subject matter for a first course in statistics. The first two units concern descriptive and exploratory data analysis, the third introduces randomness and probability, and the final three delve into statistical inference. The six units of course material are divided into smaller topics, each topic following the same structure:

- *Overview:* a brief introduction to the topic, particularly emphasizing its connection to earlier topics;
- *Objectives:* a listing of specific goals for students to achieve in the topic;
- *Preliminaries:* a series of questions designed to get students thinking about issues and applications to be studied in the topic and often to collect data on themselves;
- *In-class Activities:* the activities that guide students to learn the material for the topic;

- *Homework Activities:* the activities that test students' understanding of the material and ability to apply what they have learned in the topic;
- *Wrap-up:* a brief review of the major ideas of the topic emphasizing its connection to future topics.

In keeping with the spirit of the workshop approach, hands-on activities dominate the book. Preliminary questions and in-class activities leave enough space for students to record answers in the text itself. While comments and explanations are interspersed among the activities, these passages of exposition are purposefully less thorough than in traditional textbooks. The text contains very few solved examples, further emphasizing the idea that students construct their own knowledge of statistical ideas as they work through the activities.

While the organization of content is fairly standard, unusual features include the following:

- Probability is not treated formally but is introduced through simulations. The simulations give students an intuitive sense of random variation and the idea that probability represents the proportion of times that something would happen in the long run. Because students often have trouble connecting the computer simulation with the underlying process that it models, the text first asks students to perform physical simulations involving dice and candies to help them understand the process being modeled.
- The Central Limit Theorem and the reasoning of statistical inference are introduced in the context of a population *proportion* rather than a population *mean.* A population proportion summarizes all of the relevant information about the population of a binary variable, allowing students to concentrate more easily on the concepts of sampling distribution, confidence, and significance. These ideas are introduced through physical and computer simulations which are easier to conduct with binary variables than with measurement variables. Dealing with binary variables also eliminates the need to consider issues such as the underlying shape of the population distribution and the choice of an appropriate parameter.
- Exploratory data analysis and data production issues are emphasized throughout, even in the units covering statistical inference. Most activities that call for the application of inference procedures first ask students to conduct an exploratory analysis of the data; these analyses often reveal much that the inference procedures do not. These activities also guide students to question the design of the study before drawing conclusions from the inference results. Examples used early in the text to illustrate Simpson's paradox and biased sampling reappear in the context of inference, reminding students to be cautious when drawing conclusions.

Acknowledgments

I am privileged to teach at Dickinson College, where I enjoy an ideal atmosphere for experimenting with innovative pedagogical strategies and curriculum development. I thank my many colleagues and students who have helped me in writing this book.

Nancy Baxter Hastings has directed the Workshop Mathematics Program—of which *Workshop Statistics* forms a part—with assistance from Ruth Rossow, Joanne Weissman, and Sherrill Goodlive. Barry Tesman, Jack Stodghill, Peter Martin, and Jackie Ford have taught with the book and provided valuable feedback, as have Barr von Oehsen of Piedmont College and Kevin Callahan of California State University at Hayward. Students who have contributed in many ways include Dale Usner, Kathy Reynolds, Christa Fratto, Matthew Parks, and Jennifer Becker. I also thank Dean George Allan for his leadership in establishing the productive teaching/learning environment that I enjoy at Dickinson.

I appreciate the support given to the Workshop Mathematics Program by the Fund for the Improvement of Post-Secondary Education, the U.S. Department of Education, and the National Science Foundation, as well as by Dickinson College. I also thank Springer-Verlag for their support, particularly Jerry Lyons, Liesl Gibson, and Steven Pisano. I thank Sara Buchan for help with proofreading.

Much of what I have learned about statistics education has been shaped by the writings from which I quote above. I especially thank Joan Garfield, George Cobb, Tom Short, and Joel Greenhouse for many enlightening conversations.

Finally, I thank my wonderful wife Eileen, without whose support and encouragement I would not have completed this work. Thanks also to my feline friends Eponine and Cosette.

ALLAN J. ROSSMAN
Dickinson College
December 1995

References

Cobb, George W. (1987), "Introductory Textbooks: A Framework for Evaluation." *Journal of the American Statistical Association*, 82, 321–339.

Cobb, George W. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. Lynn Steen, MAA Notes Number 22, 3–43.

Garfield, Joan (1995), "How Students Learn Statistics," *International Statistical Review*, 63, 25–34.

Moore, David S. (1992), "Teaching Statistics as a Respectable Subject," in *Statistics for the Twenty-First Century*, eds. Florence and Sheldon Gordon, MAA Notes Number 26, 14–25.

Moore, Thomas L., and Rosemary A. Roberts (1989), "Statistics at Liberal Arts Colleges," *The American Statistician*, 43, 80–85.

To the Instructor

We want to emphasize from the outset that there is no one "right" way to teach with this book. We hope that it will prove useful to students and instructors in a wide variety of settings. Naturally, we think that the text will work best in a classroom environment that promotes the features extolled in the preface: active learning, conceptual understanding, genuine data, and use of technology.

The following suggestions are based on our own experiences and on those of many instructors who have taught with the original version of *Workshop Statistics:*

1. Take control of the course.

While this may seem obvious, we feel the "control" needed in the course differs from the traditional lecture setting but is still quite important. Students need to see that the instructor is monitoring and facilitating the progress of the course and that there is a pedagogical purpose behind all of the classroom activities.

2. Keep the class roughly together.

Part of the control that needs to be taken is to keep the students roughly together with the material, not letting some groups get too far ahead while others lag far behind.

3. Allow students to discover.

We encourage you to resist the temptation to tell students too much. Rather, let them discover the ideas and conduct analyses for themselves, while you point them in the right direction as needed. This principle of self-discovery enables students to construct their own knowledge, ideally leading to a deeper understanding of fundamental ideas and a heightened ability to apply these ideas beyond this course.

4. Promote collaborative learning among students.

This course provides a natural occasion for encouraging students to work in groups, allowing them to collaborate and learn from each other as well from you and the book. We have found such collaboration to be an invaluable part of the learning experience.

5. Encourage students' guessing and development of intuition.

We believe that much can be gained by asking students to think and make predictions about issues and data before detailed analysis. We urge you to

give students time to think about and respond to "Preliminaries" questions in the hope that these questions lead students to care more about the data they will analyze, as well as to gradually develop their own statistical intuition.

6. Lecture when appropriate.

By no means do we propose that you never speak to the class as a whole. In many circumstances interrupting the class for a "mini-lecture" is appropriate and important. As a general rule, though, we advocate lecturing on an idea only after students have begun to grapple with it first themselves.

7. Have students do some work by hand.

While we believe strongly in using technology to explore statistical phenomena as well as to analyze genuine data, we think that students have much to gain by first performing small-scale analyses by hand. We feel particularly strongly about this point in the context of simulations, where students can better comprehend the process of simulation through physical examples before proceeding to computer simulations. We also encourage instructors to assign a mixture of problems to be solved by hand and with the computer.

8. Use technology as a tool.

The counterbalance to the previous suggestion is that students should come to regard technology as an invaluable tool both for analyzing data and for studying statistics. After you have given students the chance to do some small-scale displays and calculations by hand, we suggest that you then encourage them to use technology to alleviate their computational burdens.

9. Be pro-active in approaching students.

As your students work through the activities, we strongly suggest that you not wait for them to approach you with questions. Rather, approach them to check their work and provide quick feedback.

10. Give students access to "right" answers.

Some students are fearful of a self-discovery approach because they worry about discovering the "wrong" things. We appreciate this objection and feel it makes a strong case for giving students regular and consistent feedback, including access to model answers.

11. Provide plenty of feedback.

This suggestion closely follows the two previous ones about being pro-active and providing "right" answers. An instructor can supply much more personalized, in-class feedback with this "workshop" approach than in a traditional lecture setting.

12. Stress good writing.

We regard writing-to-learn as an important aspect of *Workshop Statistics,* although it is certainly a feature that students resist. Many activities call for students to write interpretations and explanations of their findings, and we urge you to insist that students relate these to the context at hand.

13. Implore students to read well.

Students can do themselves a great service by taking their time and reading carefully. By reading directions and questions well, students can better know what is expected in an activity. Moreover, the book's expository passages interspersed among the activities contain a great deal of information that is essential for students to understand.

14. Have fun!

We sincerely hope that you and your students will enjoy a dynamic and productive learning environment as you study with *Workshop Statistics.*

15. Make use of our web resources.

Our Web sites contain a wide variety of resources ranging from downloadable worksheets and macros to an instructor's guide to sample exams to updated data files. They also contain links to sites of other instructors using *Workshop Statistics.* They can be found at:

http: //www.dickinson.edu/~rossman/ws/index.html
http: //www.uop.edu/cop/math/chance/ws/index.html

To the Student

We hope that you will find statistics to be both an important and an engaging subject to study. We want you to be aware of three principles that guided our writing of this book; the first two relate to the study of statistics generally and the third pertains to the distinctive nature of this book:

1. Statistics is not number-crunching.

Contrary to its popular perception, statistics involves much more than numerical computations. In this book you will be asked to concentrate on understanding statistical concepts and on interpreting and communicating the results of statistical analyses. In other words, you will be expected to learn to construct and analyze *numerical arguments*, using data to support your statements. In contrast to most mathematics courses, you will be using phrases such as "there is strong evidence that . . . " and "the data suggest that . . . " rather than "the exact answer is . . . " and "it is therefore proven that . . . ." In order to allow you to better focus on this understanding and communication, you will use the computer program Minitab to alleviate computational drudgery. Minitab will also present the ideas in a visual, interactive environment, allowing you to more easily understand the concepts and their properties.

2. Statistics involves the analysis of genuine data.

Supporting our contention that statistics is applicable in everyday life and in most fields of academic endeavor, you will analyze genuine data from a wide variety of applications throughout the course. Many of these data sets involve information that you will collect about yourselves and your peers; others will come from sources such as almanacs, journals, magazines, newspapers, and books. We hope that by spanning a wide variety of subject matter, the contexts will be of interest to a general audience.

3. Understanding results from investigation and discovery.

The structure of this text asks you to spend most of your time actively engaged with the material as opposed to passively taking notes. The activities have been carefully designed and tested to lead you to discover fundamental statistical ideas for yourself, in collaboration with your peers, your instructor, and the computer.

You should try to read very carefully. Pay particular attention to expository passages of the text, which summarize the most important concepts and principles, to supplement the knowledge that you construct.

Our advice to you for success in this course can be summed up in two words: *think* and *participate*. This course will ask you to think critically and to defend your arguments. Moreover, you will be asked to make guesses and collect data and draw conclusions and write summaries and discuss findings and explore alternatives and investigate scenarios and. . . . You must have an open and active mind in order to complete these tasks; in other words, you must accept responsibility for your own learning. Our responsibility as authors has been to provide you with a resource that will facilitate this learning process and lead you on the path toward understanding statistics.

Two final words of advice: *Have fun!* We sincerely hope that you will enjoy a dynamic and interactive learning environment as you study statistics.

Contents