THE WORKSHOP MATHEMATICS PROJECT

WORKSHOP STATISTICS

DISCOVERY WITH DATA

ALLAN J. ROSSMAN

Workshop Statistics

Discovery with Data

Allan J. Rossman
DICKINSON COLLEGE

Springer

Library of Congress Cataloging-in-Publication Data
Rossman, Allan J.
Workshop statistics : discovery with data / Allan J. Rossman.
p. cm.
Includes bibliographical references.
ISBN 0-387-94497-4 (softcover : alk. paper)
1. Statistics. I. Title. II. Series.
QA276.12.R673 1995
001.4'22—dc20 95-16797

Printed on acid-free paper.

Production managed by Steven Pisano; manufacturing supervised by Joe Quatela.
Photocomposed by University Graphics, Inc., 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 0-387-94497-4 Springer-Verlag New York Berlin Heidelberg

Workshop Statistics

Springer
New York
Berlin
Heidelberg
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Santa Clara
Singapore
Tokyo

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 activities 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

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.

Contents

List of Activities

NOTE: In-Class Activities appear in **boldface**.