

THE ACTIVE PRACTICE OF STATISTICS

A TEXT FOR MULTIMEDIA LEARNING



DAVID S. MOORE



THE ACTIVE PRACTICE OF STATISTICS

A text for multimedia learning

David S. Moore
Purdue University



W. H. FREEMAN AND COMPANY
New York

Acquisitions editor: Holly Hodder
Associate editor: Patrick Farace
Project editor: Diane Cimino Maass
Design coordinator: Blake Logan
Cover designer: Marjory Dressler
Illustration coordinator: Bill Page
Illustrations: Publication Services
Production coordinator: Paul Rohloff
Composition: Publication Services
Manufacturing: R.R. Donnelley & Sons Company

Library of Congress Cataloging-in-Publication Data

Moore, David S.

The active practice of statistics : a textbook for multimedia learning / David S. Moore.

p. cm.

Includes index.

ISBN 0-7167-3140-1 (alk. paper)

1. Statistics—Computer-assisted instruction. I. Title.

QA276.18.M66 1997

519.5—dc21

97-21730

CIP

©1997 by W. H. Freeman and Company. All rights reserved.

No part of this book may be reproduced by any mechanical, photographic, or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted, or otherwise copied for public or private uses, without written permission from the publisher.

Printed in the United States of America

First printing, 1997



PREFACE

The Active Practice of Statistics (APS) is an introduction to statistics that emphasizes working with data and statistical ideas. It is intended to accompany the multimedia CD *ActivStats*, authored by Paul Velleman of Cornell University and published by Addison Wesley Interactive. This book is based primarily on my *Basic Practice of Statistics (BPS)* (W. H. Freeman, 1995), but with substantial changes that reflect its place as the print component of a multimedia system.

In planning a multimedia system for learning basic statistics, Paul Velleman and I quickly decided the following: (a) The modern “data and concepts” introductory course based on *BPS* or on my longer text with George McCabe, *Introduction to the Practice of Statistics* (3rd ed., W. H. Freeman, 1998) should be the intellectual basis for the multimedia system. *ActivStats* closely follows the style, content, and order of these texts. (b) A printed text is an important part of a multimedia system, both because large blocks of text belong in print rather than on a screen and because we found that students very much want a text. (c) The multimedia presentation via CD and computer should form the student’s first exposure to each lesson, with the text serving as a more portable source for reinforcement, reference, and review.

APS is therefore an unusual text. It differs from *BPS* in these ways:

- The exposition here, while complete, is briefer. I assume that readers have *ActivStats* available for a first presentation and that they will take advantage of the interactive nature of *ActivStats* to go over rough spots repeatedly.
- It is shorter in ways that go beyond the terser exposition. I have omitted material not covered by *ActivStats*. Because students are assumed to make constant use of the software in *ActivStats*, *APS* does not dwell on computational details.

- The content is customized for students who have constant access to good software. Some methods not in *BPS*, such as normal probability plots, appear in *APS* because software makes them accessible. Exercises often require software.
- There are many new examples, exercises, and data sets. Many of the data sets from this book are available on the *ActivStats* CD. For the convenience of instructors, I have attached a list of these exercises below.

Exercises contained on the *ActivStats* CD

Lesson 2	Exercises 1, 2, 4	Lesson 14	Exercises 1, 2, 3, 7, 8, 10, 12, 14
Lesson 3	Exercises 2, 4, 6, 7, 11, 13, 14, 15	Lesson 15	Exercises 1, 4, 12, 13
Lesson 4	Exercises 1, 5, 11, 12, 14, 17, 18, 19, 20	Lesson 16	Exercises 3, 4, 7, 8
Lesson 5	Exercises 2, 3, 5, 7, 8, 9, 11, 12, 13, 17	Lesson 17	Exercises 4, 7, 9, 10, 13, 14
Lesson 6	Exercises 2, 3, 4, 7	Lesson 18	Exercises 5, 6, 7, 9, 12
Lesson 7	Exercises 1, 2, 3, 4, 5, 8	Lesson 19	Exercises 3, 4, 5, 7, 10, 11, 13, 14, 16, 18
Lesson 8	Exercises 2, 3, 4, 5, 6	Lesson 20	Exercises 4, 6, 7, 10, 12, 13, 14, 15
Lesson 9	Exercises 1, 2, 9, 10, 12, 14, 15	Lesson 21	Exercises 2, 6, 8, 9, 10, 13, 16
Lesson 10	Exercises 3, 4, 9, 11, 15	Lesson 22	Exercises 1, 2, 5, 9, 10, 15
Lesson 11	Exercises 3, 5, 7, 10, 14	Lesson 23	Exercises 1, 2, 6, 8, 12, 13
Lesson 12	Exercises 1, 2, 7, 8, 9, 10	Lesson 24	Exercises 1, 2, 3, 4, 5, 10
Lesson 13	Exercises 1, 3, 8, 9		

The lesson and part structure of this book mirrors that of *ActivStats*. For each of the six parts, I have added review lessons that offer students an outline-style review and a set of review exercises. Lesson 1 is a brief discussion of how to use *APS* along with *ActivStats*. You may wish to read that lesson for more comments on the interaction between CD and book, and to read the Introduction for a brief statement of the approach to basic statistics that both the CD and the book encourage.

What about those other books?

This text shares the basic spirit of my longer texts, *The Basic Practice of Statistics* (*BPS*) and *Introduction to the Practice of Statistics* (*IPS*, with George McCabe). It offers a modern presentation of statistics that gives proper emphasis to analysis of data and design of data production in addition to formal probability-based inference, stresses the practical and conceptual links

between these aspects of the subject, and concentrates on core concepts and strategies that carry over into more advanced statistical settings. Most of all, the exposition centers on understanding what data tell us about substantive problems. *APS* of course contains many new examples, exercises, and data sets. It also differs from the other texts in systematic ways.

The intent of *BPS* is to present the core concepts and methods of data-oriented statistics in a manner accessible to relatively non-quantitative students. *BPS* does not assume that students will use technology beyond a two-variable statistics calculator, and it deliberately reduces the usual presentation of probability in order to concentrate on core statistics. *IPS* is a longer book that presents some more advanced statistical material as well as a quite traditional treatment of elementary probability. Its discussions have greater depth and richness; this is a substantial bonus for some groups of students, but a barrier for others.

The efficiency of good software allows *APS* to present some material not present in *BPS* without raising barriers to students. Normal probability plots appear throughout. The presentation of probability is much more complete than in *BPS*; in particular, there is a careful introduction to conditional probability. On the other hand, the topics discussed in *APS* are strictly limited to those present in *ActivStats*. *BPS* adds an introduction to oneway analysis of variance. *IPS* goes on to present multiple comparisons, twoway ANOVA, and some basics of multiple regression.

It has been a pleasure to work with Paul Velleman in thinking about multimedia instruction, and I thank him for his careful reading of this manuscript. We have aimed at close coordination between CD and book, but readers should be aware that software can be revised much more quickly than a text. It has once again been a pleasure to work with W. H. Freeman's excellent staff, especially project editor Diane Cimino Maass; designer, Blake Logan; production coordinator, Paul Rohloff; and copy editor Pamela Bruton.

I very much hope that *The Active Practice of Statistics*, in company with *ActivStats*, will make a contribution to moving statistics instruction closer to statistical practice in an era when the automation of calculations makes data judgment and understanding of statistical principles more important than ever.

David S. Moore



INTRODUCTION: WHAT IS STATISTICS?

Why study statistics?

Statistics is the science of gaining information from numerical data. We study statistics because the use of data has become ever more common in a growing number of professions, in public policy, and in everyday life. Here are some examples of statistical questions raised by a viewer of the nightly news.

- The Bureau of Labor Statistics reports that the unemployment rate last month was 6.5%. The government didn't ask me if I was unemployed. How did they obtain this information? How accurate is that 6.5%?
- Another news item describes restrictions on smoking in public places. I hear that much of the evidence that links smoking to lung cancer and other health problems is "statistical." What kind of evidence is statistical evidence?

- A medical reporter cites a study claiming to show that regular physical exercise leads to longer life. But a doctor she interviews casts doubt on the study's usefulness. How can I tell whether the data from a study really support the conclusions that are announced?
- Here's a special report on the international competitiveness of American industry. The experts interviewed talk about improving quality and productivity through better management, new technology, and effective use of statistics. What can statistics do except keep score?

We can no more escape data than we can avoid the use of words. Like words, data do not interpret themselves but must be read with understanding. Just as a writer can arrange words into convincing arguments or incoherent nonsense, so data can be convincing, misleading, or just irrelevant. Numerical literacy, the ability to follow and understand arguments based on data, is important for everyone. The study of statistics is an essential part of a sound education.

What is statistics?

For most users of statistics, and even for most professional statisticians, statistics provides tools and ideas for using data to gain understanding of some other subject. Statistics in practice is applied to study the effectiveness of medical treatments, the reaction of consumers to television advertising, the attitudes of young people toward sex and marriage, and much else. Although statistics has an impressive mathematical theory, we are concerned with the *practice* of statistics. We can divide statistics in practice into three parts:

1. **Data analysis** concerns methods and ideas for organizing and describing data using graphs, numerical summaries, and more elaborate mathematical descriptions. The computer revolution has brought analysis of data back to the center of statistical practice. Statisticians have responded with new tools and (more important) new organizing ideas for exploring data. Parts I and II of this book discuss data analysis.
2. **Data production** supplies methods for producing data that can give clear answers to specific questions. Basic concepts about how to select samples and design experiments are perhaps the most influential of all statistical ideas. These concepts are the subject of Part III.

3. **Statistical inference** moves beyond the data in hand to draw conclusions about some wider universe. Statistical inference not only draws conclusions but accompanies those conclusions with a statement about how trustworthy they are. Inference uses the language of probability, which we introduce in Part IV. Because we are concerned with practice rather than theory, we can function with a quite limited knowledge of probability. Part V discusses the reasoning of statistical inference and methods for practical inference in several simple settings. Part VI offers introductions to inference in some more complex settings.

How will we approach the study of statistics?

The goal of statistics is to gain understanding from data. Data are numbers, but they are not “just numbers.” *Data are numbers with a context.* The number 10.5, for example, carries no information by itself. But if we hear that a friend’s new baby weighed 10.5 pounds at birth, we congratulate her on the healthy size of the child. The context engages our background knowledge and allows us to make judgments. We know that a baby weighing 10.5 pounds is quite large, and that it isn’t possible for a human baby to weigh 10.5 ounces or 10.5 kilograms. The context makes the number informative.

Because data are numbers with a context, doing statistics means more than manipulating numbers. This book is full of data, and each set of data has some brief background to help you understand what the data say. Examples and exercises usually 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 very important to form the habit of asking “What do the data tell me?” rather than just concentrating on making graphs and doing calculations. This book tries to encourage good habits.

Nonetheless, statistics involves lots of calculating and graphing. You will use computer software to automate most calculations and graphs. Because graphing and calculating are automated in statistical practice, the most important assets you can gain from the study of statistics are an understanding of the big ideas and the beginnings of good judgment in working with data. Ideas and judgment can’t (at least yet) be automated. They guide you in telling the computer what to do and in interpreting its output. This book tries to explain the most important ideas of statistics, not just teach

methods. Some examples of big ideas that you will meet (one from each of the three areas of statistics) are “always plot your data,” “randomized comparative experiments,” and “statistical significance.”

These, then, are the principles that should guide your learning of statistics:

- Try to understand what data say in each specific context. All the methods you will learn are just tools to help understand data.
- Let a computer do as much of the calculating and graphing as possible, so that you can concentrate on what to do and why.
- Focus on the big ideas of statistics, not just on rules and recipes.

But perhaps the basic principle of all learning is persistence. The main ideas of statistics, like the main ideas of any important subject, took a long time to discover and take some time to master. The gain will be worth the pain.

CONTENTS

Preface	ix
Introduction: What Is Statistics?	xii
LESSON 1 How to Use This Book	1

PART I Understanding Data 5







LESSON 2 Data and Measurement	6
LESSON 3 The Distribution of One Variable	10
LESSON 4 Measuring Center and Spread	27
LESSON 5 Normal Distributions	43
Part I Review	59

PART II Understanding Relationships 69



LESSON 6 Comparing Groups	70
LESSON 7 Scatterplots	77
LESSON 8 Correlation	84
LESSON 9 Least-Squares Regression	92
Part II Review	108

PART III	Generating Data	117
	LESSON 10 Sample Surveys	118
	LESSON 11 Designed Experiments	129
	Part III Review	139
PART IV	Experience with Random Behavior	145
	LESSON 12 Randomness	146
	LESSON 13 Intuitive Probability	152
	LESSON 14 Conditional Probability (optional)	163
	LESSON 15 Random Variables	174
	LESSON 16 Sampling Distributions	184
	Part IV Review	197
PART V	Statistical Inference	205
	LESSON 17 Estimating with Confidence	206
	LESSON 18 Confidence Intervals for a Mean	222
	LESSON 19 Testing Hypotheses	233
	LESSON 20 Tests for a Mean	248
	Part V Review	263
PART VI	Topics in Inference	271
	LESSON 21 Comparing Two Means	272
	LESSON 22 Inference for Proportions	290
	LESSON 23 Two-Way Tables	311

LESSON 24	Inference for Regression	329
	Part VI Review	351
Notes and Data Sources		367
Appendix		
Table A	Standard Normal Probabilities	376
Table B	The t Distributions	378
Table C	The Chi-Square Distributions	379
Solutions to Selected Exercises		381
Data Table Index		428
Index		429



LESSON 1

HOW TO USE THIS BOOK

The Active Practice of Statistics (APS) is intended to be used in company with the multimedia *ActivStats* system. It is therefore somewhat different from other texts. *ActivStats* has important advantages as a tool for learning. In particular, you interact with the ideas and examples presented as you move through a lesson in *ActivStats*. The active learning encouraged by a good multimedia system is a better starting point than reading the same material in a book.

We recommend that you begin your study of each lesson with *ActivStats*. Then turn to this book for reinforcement and a record of the lesson content.

The organization, notation, and basic content of *APS* exactly match *ActivStats*. Each of the 24 lessons in this book corresponds to an *ActivStats* lesson. Many of the exercises in this book (especially those accompanied by sets of data) are available within *ActivStats* through the homework icon

in the tool bar of each lesson. In addition, *APS* offers a review outline of each part of *ActivStats*, with lists of the specific skills you should have acquired from each part and review exercises to test those skills.

The presentation of the content of each lesson in this book is quite brief. It is intended to give you a compact statement that contains all the essentials but assumes that you have made a first acquaintance with the material during your work on the *ActivStats* lesson. Seeing two different expositions (multimedia and print) will solidify your understanding. You will also find discussions of some ideas that do not lend themselves to multimedia presentation.

SOFTWARE TOOLS

In practice, users of statistics always use software to automate their calculations and graphics. This book fits actual statistical practice in that it assumes that you will use software in almost all your work. *ActivStats* makes available to you software tools for doing the many calculations and graphs that modern statistics requires. These are of two kinds:

- The *Data Desk* statistics software, a modern menu-driven statistics software package. *Data Desk* is widely used for statistical analyses. Moreover, its organization is generally similar to that of other software for doing statistics, so that experience with *Data Desk* equips you to do statistical work in practice.
- Animated graphics tools that make it easy to do some things that do not commonly appear in statistics software. These tools are used in *ActivStats* to help you learn, and they are available in the lesson tool bars for you to use detached from the lessons.

This book assumes that you will use the software tools made available in *ActivStats*. Exercises require use of these or similar tools. The text does not tell you how to do calculations and graphs by hand except where knowing how helps your understanding.

You will find that *ActivStats* itself will teach you how to use the *Data Desk* software and the animated tools. Because you should start with the *ActivStats* lessons before reading the corresponding lessons in this book,

there is no instruction here on how to use the software. One benefit of doing the exercises is to ensure that you can use the software effectively outside the *ActivStats* lessons that teach you how to use it.

WHY DO I NEED A TEXTBOOK?

Multimedia systems are in many respects more effective than print as aids to learning. But printed books retain some important advantages:

- Large blocks of text are much easier to read and work with in print than on the screen. *APS* complements *ActivStats* by offering detailed explanations, worked examples, and comments on statistical practice that are better presented in a book.
- A book is more portable than a CD because you don't need a computer on hand. You can study *APS* anywhere.
- Because of its portability and because it is easy to find your way around in a book, a permanent record of what you must know is more convenient in book form.

Think of the lessons in *APS* as detailed and permanent records of the content of the *ActivStats* lessons, lacking only software instruction.



FLORENCE NIGHTINGALE

Florence Nightingale (1820-1910) won fame as a founder of the nursing profession and as a reformer of health care. As chief nurse for the British army during the Crimean War, from 1854 to 1856, she found that lack of sanitation and disease killed large numbers of soldiers hospitalized by wounds. Her reforms reduced the death rate at her military hospital from 42.7% to 2.2%, and she returned from the war famous. She at once began a fight to reform the entire military

health care system, with considerable success.

One of the chief weapons Florence Nightingale used in her efforts was data. She had the facts, because she reformed record keeping as well as medical care. She was a pioneer in using graphs to present data in a vivid form that even generals and members of Parliament could understand. Her inventive graphs are a landmark in the growth of the new science of statistics. She considered statistics essential to understanding any social issue and tried to introduce the study of statistics into higher education.

In beginning our study of statistics, we will follow Florence Nightingale's lead. Part I stresses the analysis of data as a path to understanding. Like her, we will start by measuring outcomes and making graphs to see what data can teach us. Along with the graphs we will present numerical summaries, just as Florence Nightingale calculated detailed death rates and other summaries. Data for Florence Nightingale were not dry or abstract, because they showed her, and helped her show others, how to save lives. That remains true today.