BEGINNING STATISTICS A TO Z

WILLIAM MENDENHALL

DISK ENCLOSED

COMPLETE SET OF MEMORY CARDS INCLUDED WITH EVERY NEW COPY OF BOOK

Beginning Statistics: A to Z

William Mendenhall
University of Florida, Emeritus

An Alexander Kugushev book

Duxbury Press
An Imprint of Wadsworth Publishing Company
Belmont, California

Print Buyer: Karen Hunt
Designer: Nancy Benedict Graphic Design
Technical Illustrator: Lori Heckelman
Cover: Vargas/Williams/Design
Cover Background: West Light
Compositor: Polyglot Compositors, Pte., Ltd.; GTS Graphics, Inc.
Printer: R. R. Donnelley

This book is printed on acid-free recycled paper.

Duxbury Press
An Imprint of Wadsworth Publishing Company
A division of Wadsworth, Inc.

Printed in the United States of America

1 2 3 4 5 6 7 8 9 10—97 96 95 94 93

Library of Congress Cataloging-in-Publication Data

Mendenhall, William.
Beginning statistics—A to Z / William Mendenhall.
p. cm.
Includes index.
ISBN 0-534-19122-3
1. Statistics. I. Title. II. Title: Beginning statistics.
QA276.12.M45 1993
51925—dc20

92-31635
CIP

Beginning Statistics: A to Z

preface

Most of us would like our students in an introductory course to acquire a usable understanding of statistics that stays with them one, two, or five years (or even longer) after they have completed the course. This may be an impossible dream but a change in emphasis may bring us closer to this goal.

Do we need to spend so much time on the basic theory of probability? Does the student need to grind and grind through numerous calculations when, in this day of the computer, few, if any, calculations are performed on a hand calculator?

A student should understand the role that probability plays in statistical inference. But how much probability theory does a student really need to know in order to understand these concepts? Calculating our way through a simple example to better understand why and how a statistic works is useful, but maybe one is enough. The time that we currently spend on probability and calculating will be better spent on statistical concepts, learning about some of the most useful statistical methods, how and why they work, and, most important, what they mean.

This text is designed to focus on **concepts, statistical analyses,** and their **interpretation.** Downplaying probability and calculating increases the time available for these topics. It also allows us to increase course coverage. The book presents a complete introduction to statistics and statistical methods in one term. We would expect courses to include the standard topics contained in Chapters 2–7 along with all or most of the methods chapters that follow.

A Comment on Probability

We use the concept of the "rare" (i.e., improbable) outcome throughout the text to explain the reasoning employed in statistical inference. The student learns in Chapter 3 that it is very improbable that an observation will fall more than three standard deviations from its mean and we use this concept to explain how control charts are used in process quality control. Similarly, sampling distributions are used to identify rare or improbable values of statistics and enable us to establish upper limits on errors of estimation (what the news media call sampling error). We apply the same reasoning when we base a decision in a statistical test on the p-value of the observed test statistic. These difficult inferential concepts are introduced in Chapter 5 and reinforced repeatedly in the methods chapters that follow.

A Comment on the Use of Data

Most examples and exercises are based on data selected from newspapers or research journals. These data add realism to the presentation and, more important, provide evi-

dence that a knowledge of statistics is a very helpful tool for most professions. They also provide realistic data sets for instructor-generated examples and exercises.

A computer printout containing the relevant statistics for each confidence interval or statistical test is shown and interpreted. The student is also shown how to calculate pertinent statistics because many journal articles include only sample descriptive statistics. These examples and exercises can be omitted to further reduce calculations if the instructor wishes to do so.

A Comment on Flash Cards

We think with words and concepts! True, we can learn the words and concepts in the process of "doing"—i.e., solving problems, etc.—but it won't be as fast or as efficient as memory reinforcement *along with* the "doing." To this end, the student is supplied with "flash (memory) cards" that are shrink-wrapped with the book. They are accompanied by an explanation on their use.

Supporting Materials

The following supporting material is available for the instructor:

1. An instructor's manual that contains suggestions by the author on the use of the text and solutions to the exercises
2. A floppy disk that contains the three data sets listed in Appendix 1 of the text
3. A teaching disk that can be used to demonstrate statistical concepts (e.g., the concept of a sampling distribution, the Central Limit Theorem, etc.)
4. An examination bank

The author acknowledges with thanks the assistance of many talented people. Ken Brown (College of San Mateo), Carolyn Apperson Hansen (University of Virginia Health Sciences Center), Laurence Johnson (Cuesta College), James Lang (Valencia Community College), Austin Meek (Canada College), Kirk Steinhorst (University of Idaho), and Bruce Trumbo (California State University, Hayward) provided early reviews of the manuscript and helpful suggestions for its improvement. Dennis Wackerly of the University of Florida provided many suggestions over lunch and also reviewed Chapter 11. One of his most useful suggestions was directing me to James Lang of Valencia Community College to obtain the excellent instructional disk that accompanies the text. I particularly thank Carolyn Hansen for producing the computer printouts, for finding the cholesterol data set, and for suggestions throughout the writing. I am grateful to Terry Sincich, J. J. Cerda, M.D., journal editors, and others for permission to use data sets and tables, and to Teresa Bittner of Laurel Tutoring for preparing the solutions manual and answers to exercises.

I thank Alex Kugushev, long-time editor and friend, for reading and critiquing early chapters of the manuscript and helping to improve my writing style. And, particularly, I acknowledge the outstanding and meticulous copyediting and comments of Susan Reiland. Finally, I thank my wife for her support and for tolerating my writing "just one more book."

table of contents

1 2 3 4 5 6 7 8 9 10 11

one

To the Reader

1.1 The Role of Statistics in the Modern World

We live in a world described by numbers—numbers that monitor every aspect of our daily lives and of the world in which we live. The data to support this contention? See your daily newspaper, news magazine, or one of the many professional or scientific journals. The varied applications of statistics can be seen in the following examples.

Manufacturing

The Japanese did not achieve superiority over some segments of American industry solely on the basis of labor costs. They surpassed us on quality. To do this, they collected data on the quality characteristics of a product, say a Toyota automobile. The number of breakdowns per 100,000 miles, the causes of breakdowns, and many other data on the current quality of a model were recorded. Japanese engineers had to find out what the quality was today. Then, from an analysis of the data, they were able to make a better product for tomorrow. The tools used to deduce the necessary changes? The techniques of W. Edwards Deming, the world recognized statistician and expert on statistical quality control—and statistics.

The Social Sciences

An article in *The Gainesville Sun* (October 26, 1991) is headlined "Study Indicates Bias in Bar Exam." It goes on to say that Hari Swaminathan, a statistician at the University of Massachusetts, examined the test results of 3,777 law school graduates who took the bar exams that year. Seventy-five percent of the 3,554 white graduates passed the exams given in February and July. Only 39% of the black graduates passed the February exam; 46% passed the July tests. Is this difference in passing percentages due to random variation or is it due to a real difference in passing rates of whites and blacks? If a real difference exists in the passing rates, can we say that it is due to "bias"? What can statistics tell us about this complex question?

Business

In the fall of 1991, the economic picture looked bad to some economists. Why? The statistics that characterize the state of the economy—the unemployment rate, the measure of consumer confidence, new housing construction starts, consumer spending, etc.—were all headed in the wrong direction. How do we monitor the state of the economy? Statistics.

Politics

The New York Times (July 14, 1991) reports that, "Poll Finds G.O.P. Growth Erodes Dominant Role of the Democrats." The article indicates that of the 14,685 persons polled, 34% called themselves Democrats and 31% called themselves Republicans. These statistics, numbers that summarize and describe the political pulse of the nation, tell us the party affiliation of voters at the time the poll was taken. But the pulse changes from day to day. As an election draws near, polls are taken more and more frequently to monitor changes. The tool used to evaluate the changes and to deduce their implications? Statistics.

Research

A headline in *The Wall Street Journal* (November 5, 1991) states, "Pfizer Gets Clearance from FDA to Market Zithromax Antibiotic." This announcement of approval by the U.S. Food and Drug Administration (FDA) to market a new and powerful antibiotic carries an unspoken message. Behind the approval, as for all new food and pharmaceutical products, lie years of testing, by both the manufacturer and the FDA. The object is to prove that the new product works and to establish safe dosages for the consumer. The method used to draw these conclusions from experimental data? Statistics.

These examples illustrate just a few of the many applications of statistics. Most important, they show that statistics enters into almost all aspects of human endeavor. Why?

Statistics is about describing and predicting. It is a response to the limitations of our ability to know. Any complex aspect of human knowledge is either difficult for us to understand, or lends itself to misunderstanding—unless we possess the right tool with which to examine and analyze its complexity. Statistics is that tool. You will need it as you enter the 21st century.

1.2 The Language of Statistics

How does statistics differ from many other courses that you have taken, or will take, in college? Obviously, the subject matter is different, but it is more than that. The material in most courses in the social and biological sciences, business, etc., is presented in everyday language. You encounter new words and concepts, of course, but you have had some prior exposure to these subjects in reading news or magazine articles.

Statistics, in contrast, will seem like a new language. True, you have read articles similar to those described in Section 1.1. But it is likely that you have had minimal

exposure to the language and concepts of statistics. Therefore, you need to develop a new attitude and be prepared to learn a new language.

We think with words and concepts, and one topic builds on another. **In order to follow your instructor's lectures and to read this book, you need to commit new words, symbols, formulas, and concepts to memory. You will also need to stay with the course daily.** If you do, you will find statistics interesting—in fact, fascinating and useful.

1.3 The Role of the Computer in Statistics

Statistical methods often involve extensive, tedious numerical calculations. Prior to the invention of the computer, these calculations were done on a desk calculator. All statistical courses taught students how to use the appropriate formulas and how to simplify the calculations, if possible. Some calculations were so lengthy that it was impossible to perform them on a desk calculator.

The advent of the computer solved this problem. It is still necessary to know what a statistical quantity is and to understand why and how it works. But, in this day and age, statistical calculations are done on a computer. Consequently, we will downplay calculating. Instead, we will learn how to read and interpret statistical computer printouts and published statistical statements. If you have access to a computer and a statistical software package, you may want to try producing your own output for a statistical analysis of some data.

There are a number of good statistical software packages available. The printouts for these packages differ in minor respects. Once you are familiar with one, you can easily learn how to read another. We will show primarily printouts for the Minitab, SAS, and Data Desk statistical software packages with a few others used for exercises.*

Some statistical software packages are better than others and some are easier to use. If you want to perform some statistical analyses on your computer, ask your instructor to recommend a suitable software package.

1.4 How to Use This Book

This book explains the basic concepts of statistics and shows how they are applied. The material in the text will be presented in the following format:

* Minitab is a statistical software package produced by Minitab, Inc., 215 Pond Laboratory, University Park, Pennsylvania 16802.

SAS is a statistical software package produced by the SAS Institute, Box 8000, Cary, North Carolina 27512.

Data Desk is a computer software package developed by Paul Velleman, Data Description, Inc., Box 4555, Ithaca, New York 14852.

1. Each chapter begins with a short paragraph, entitled *In a Nutshell*, that explains how the chapter material fits into the overall objective of the text.

2. **Definitions and concepts are printed in bold type.** Key words are shown in bold italics. Formulas and symbols are shown in boxes. This is the material that you might highlight with a soft pen for review. We have done it for you. Most of the explanatory material is shown in plain text.

3. **To help you commit new words and concepts to memory, use the accompanying memory cards.** The more times you try to recall a fact, the easier it is to recall. Look at the question on the face of a memory card. Try to recall the answer. Then flip the card to the opposite side and read the answer. Use the cards for 5 or 10 minutes several times a day.

4. After we have presented new words and concepts, we will show you how to use them. Examples will show you how to apply them to real-world situations. Then we will present exercises and you can see what you have learned.

The exercise sets start with basic questions and short problems to reinforce your understanding of the language of statistics. Then we will ask you to apply the new concepts to the interpretation of computer outputs and the statistical analyses of real data sets taken from newspapers, magazines, and scientific journal articles.