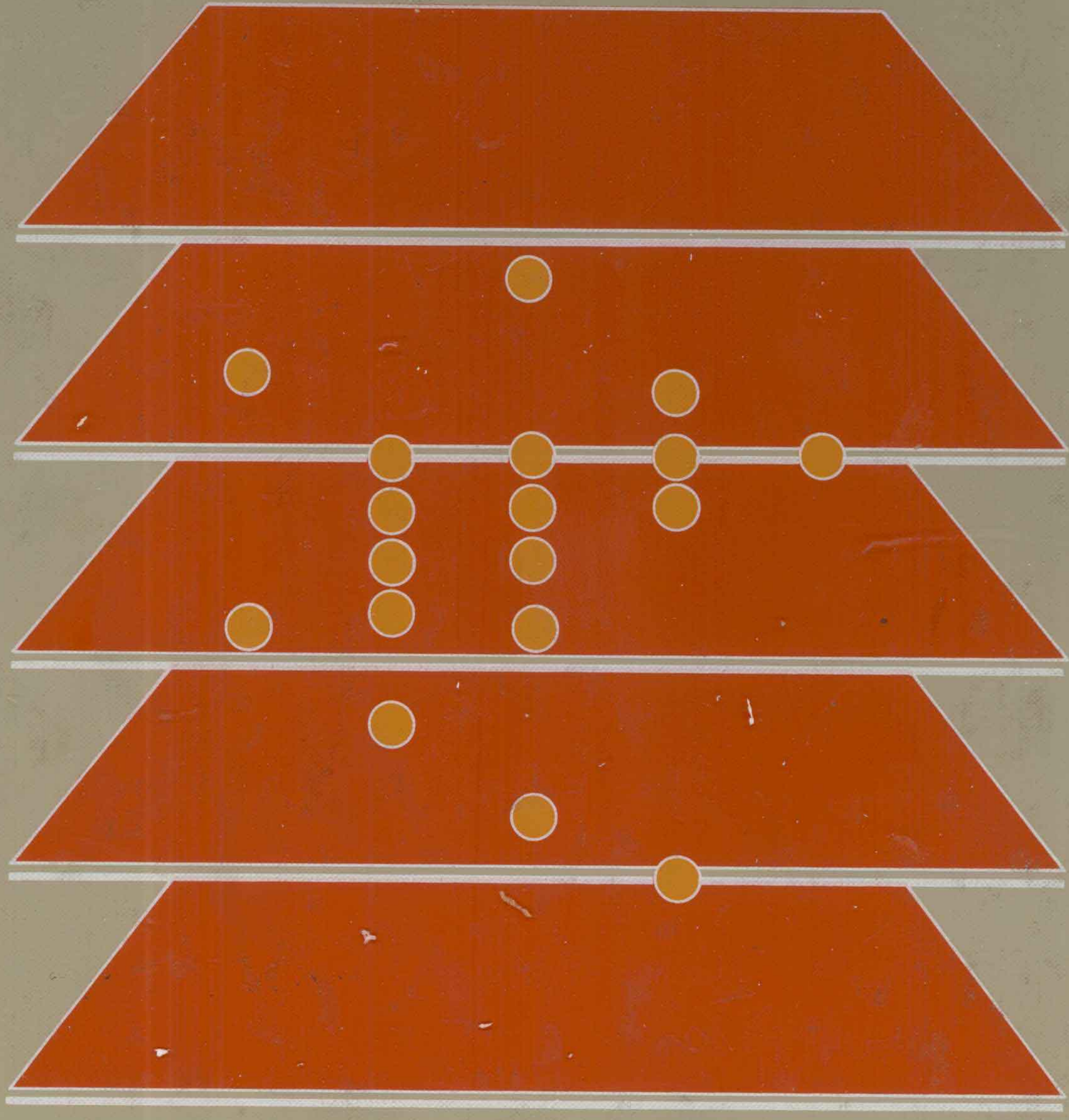
Statistics for Business and Economics



Van Matre

Gilbreath

Statistics for Business and Economics

REVISED EDITION

1983

BUSINESS PUBLICATIONS, INC.
Plano, Texas 75075

To Mindy, Kelly,
and My Parents
J.G.V.

ISBN 0-256-02885-0

Library of Congress Catalog Card No. 82–74066

Printed in the United States of America

1 2 3 4 5 6 7 8 9 0 V 0 9 8 7 6 5 4 3

Preface

Statistics for Business and Economics, revised edition, is designed primarily for undergraduate students in business, economics, and management. The introductory statistics course provides the sole exposure to statistics for many students, while other students may continue to study in this and related areas such as operations research or econometrics. To meet the needs of both groups, this book emphasizes a conceptual approach with both mathematical and verbal explanations. Selected derivations are presented, often in footnotes. Illustrations and exercises are plentiful; the amount of "busy" work is kept to a minimum.

We assume the reader's mathematical maturity is college algebra. Students with a knowledge of elementary calculus will benefit from the supplementary explanations given for certain concepts. However, calculus-based material is relegated to footnotes or optional sections so that no loss of continuity occurs if this material is omitted. Optional sections which contain calculus, selected derivations, or advanced concepts are clearly marked with asterisks.

The book may be used for courses of varying lengths. For two quarters or two semesters, we suggest that the book be covered much in its entirety. Several chapters stand alone, however, and may be considered optional. These include Chapter 3, Index Numbers; Chapter 7, Statistical Decision Theory; Chapter 14, Time Series Analysis; and Chapter 15, Nonparametric Methods. A one-quarter or one-semester course could be built around the chapters remaining after excluding these optional chapters. Chapters 9, 10, 11, and 13 may also be considered optional for these shorter courses. Note that Chapter 12, Simple Linear Regression, does *not* require any material from Chapters 10 and 11. Optional sections of chapters are marked with asterisks. (More detailed suggestions may be found in the Instructor's Manual.)

The use of computers and statistical software packages is integrated throughout the text as a natural component of statistical analysis rather than as something unique. Illustrations are presented throughout the chapters. Some exercises are designated as computer based while others are left to the discretion of the instructor or student. This approach allows the instructor complete freedom in structuring students' computer involvement. While widely available computer routines provide efficient ways to perform the more laborious calculations, we caution the reader that the use of a computer routine is not a substitute for sound statistical reasoning; it is merely a computational aid.

FEATURES OF EACH CHAPTER

Each chapter is introduced with a quotation that relates to the chapter content. Biographical sketches of major and minor figures in statistics are included to provide a historical frame of reference and to give human interest to the subject. Cartoons are used to interject humor; reprints from various sources are used to relate text discussion to current practice. Each chapter concludes with a summary of major topics. Following the chapter summary is a section that provides references to the applications appropriate to the chapter. (The Instructor's Manual also provides detailed classifications of this material, allowing instructors to use the material in a variety of ways.) Also included at the end of every chapter are convenient lists of important terms, symbols, and formulas, followed by true-false and fill-in-the blank questions, as well as a generous number of carefully designed exercises. Answers to the odd-numbered exercises are given at the end of the book.

OVERVIEW OF THE BOOK

All topics under the general heading of "descriptive statistics" are presented in a single chapter, Chapter 2. This includes frequency distributions, averages, and measures of variation. The result is a more unified and complete treatment of these closely related topics. Index numbers are presented as descriptive measures in Chapter 3. However, this topic stands alone and can be placed at any point in the sequence of topics for a given course. This material is organized so that the topic may be covered only briefly or in some depth.

Probabilistic concepts are essential for an understanding of statistical inference. A careful presentation of the basic concepts is given in Chapter 4, while probability mass functions and probability density functions are given in Chapter 5. Sampling and sampling distributions, covered in Chapter 6, complete the preparation for estimation and testing. For hypothesis testing, the modern approach to reporting the *p-value* is provided in addition to discussions of the traditional decision rules. The use of various decision rules is illustrated throughout the text.

Separate chapters are devoted to nonparametrics, multiple regression, chi-square, and analysis of variance. Chapter 7 covers decision theory through normal priors and the two-action problem without extended digressions into utility theory and the assessment of prior distributions. An expanded coverage of time series analysis, Chapter 14, includes exponential smoothing and a discussion of quadratic and exponential trend equations.

One of the unique features of this book is the devotion of a full chapter to the actual use of statistics in today's business world. Chapter 16, Statistics in Business, provides a look at how some of the basic concepts are used in accounting, banking, finance, marketing, personnel management, and other areas of application. The intent of the chapter is to help show the reader why there is a need for statistical concepts and to provide a brief look at how these concepts are used in decision making. This material can be used in at least three ways. The entire chapter, or selected applications, could be read as a motivational tool in an introduction to the course, or as each topic is introduced. Second, study of one or more methods in a given chapter could be followed by a look at a specific application. Finally, the entire chapter could serve as a capstone to the course.

SUPPLEMENTS

A comprehensive Instructor's Manual and an optional Study Guide/Workbook are available. The guide may be used independently by students as a supplement or assigned by the instructor as an integral component of the course. The Instructor's Manual provides detailed solutions to all exercises in the text as well as the solutions to the unanswered exercises in the Study Guide. The Instructor's Manual also contains suggestions on using the text in a variety of course situations. Transparency masters for selected material are included in the manual.

CHANGES IN THIS EDITION

The objectives, basic structure, and level of presentation of the revised edition remain the same as in the first edition. However, users of the first edition will find a number of minor, yet significant, changes. These include:

1. All chapters contain additional exercises.
2. The definition and use of the sample variance have been revised throughout the text to reflect the popular usage of the unbiased estimator $s^2 = \dfrac{\Sigma (x - \bar{x})^2}{n - 1}$.
3. New sections have been added to incorporate (*a*) hypothesis testing for differences in means with paired data and (*b*) the runs test in nonparametrics.
4. Expanded binomial tables now include values for $\pi > 0.5$.
5. Some notation is revised to promote easier reader comprehension.
6. Transparency masters for selected material have been added to the Instructor's Manual.
7. The Study Guide/Workbook has been expanded and updated to reflect changes in the text.

ACKNOWLEDGMENTS

To all those individuals who took the time to send comments and suggestions about the first edition, we sincerely appreciate your thoughtfulness and efforts. Be assured that all comments, regardless of their nature, were welcomed and carefully studied. Without doubt, this helped to improve the content. We also benefitted from several thoughtful reviews of the entire manuscript. While it is impossible to cite all, we must publicly thank the following individuals: Robert L. Andrews, Virginia Commonwealth University; Aaron H. Brown, Corpus Christi State University; Manfred W. Hopfe, California State University, Sacramento; Ronald S. Koot, Pennsylvania State University; David W. Pentico, Virginia Commonwealth University; Ernest M. Scheuer, California State University, Northridge; Lynette K. Solomon, Stephen F. Austin State University; Betty J. Whitten, University of Georgia; and T. Hillman Willis, Louisiana State University, Shreveport. Thanks also to M. Gene Newport, University of Alabama in Birmingham; to U.A.B. colleagues Seung-Dong Lee, Bob Stanford, and Herb Tsang; and to U.A.B. graduate assistants Ed Gillenwater and Ellen Hiller.

For typing and clerical support, we are very grateful to K. W. Hall and Jill Kramer, both of Virginia Commonwealth University, and to Kaki Culpepper, University of Alabama in Birmingham.

We are grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Longman Group Ltd., London, for permission to reprint Table III from their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th ed., 1974). To the individuals associated with the University of New Hampshire Statistics Teaching Project, we appreciate permission to reproduce and adapt material from the Statistical Test Item Collection System (STICS). We also thank the other authors and publishers cited throughout the text who allowed us to reprint various figures, tables, and excerpts from articles.

We trust our colleagues will find this book helpful in meeting their classroom needs; we welcome any comments and suggestions.

Joseph G. Van Matre **Glenn H. Gilbreath**

Contents

**Optional material.

**Optional material.

**Optional material.

**Optional material.

1 Statistics: An Overview

> . . . 'I wonder what Latitude or Longitude I've got to?' (Alice had not the slightest idea what Latitude was, or Longitude either, but she thought they were nice grand words to say).
>
> Lewis Carroll
> *Alice's Adventures in Wonderland*

1.1 INTRODUCTION

As the student begins this study of statistics, he must be apprised that many new concepts lie ahead. One can understand these concepts only by "speaking the language" of statistics. This is an accepted fundamental for an introductory course in any subject, because the particular terms involved must be mastered before they can be used. Leafing through this text, the student will observe a variety of new symbols (e.g., σ, μ) and words (e.g., median, regression). The Greek alphabet printed in Appendix A indicates how these symbols are pronounced (σ: sigma and μ: mu), but the statistical meaning of these symbols must be learned. We trust that when a student completes this book his statistical knowledge cannot be characterized by Alice's use of *longitude* and *latitude*.

1.2 STATISTICS DEFINED

The word *statistics* calls forth different things to different people. For example, statistics are Tony Dorsett's yards per carry and season touchdowns, or the price-earnings ratio of a stock. However, as we shall see, the study of statistics possesses great utility in capacities other than providing a means of description, although this aspect should not be ignored.

A somewhat traditional definition of the discipline is:

> *statistics:* a body of methods dealing with the collection, description, analysis, and interpretation of information that can be given in numerical form.

Since the definition above is rather broad, we consider the subject in more detail by discussing the major divisions within the field; namely, descriptive statistics, inferential statistics, and statistical decision theory.

Descriptive Statistics

Descriptive statistics are measures that are used to characterize (describe) a mass of numerical data. For example, suppose one were given the SAT scores of the 300 freshmen of the School of Business of some university and then later were asked, "How do the B-school students' entrance scores look?" Probably, no one would reply by beginning to read the 300 scores. An obvious possible descriptor is what the statistical layman calls the average. This and other descriptive measures will be the subject of the next chapter. Art Buchwald's column of September 16, 1971, in the *Washington Post* provides another illustration on the use (or misuse?) of descriptive statistics:

There was good news out of Washington last week. According to Attorney General John Mitchell, President Nixon's war on crime has been successful, and the results of the administration's monumental efforts have been so great that "fear is being swept from the streets of some—though not all—American cities." . . .

The reason for the euphoria in the Justice Department is that FBI statistics for 1970 indicated that the rate of increase of crime had gone down from 12 percent in 1969 to 11.3 percent in 1970.

This sounded terrific, until I read that the same statistics revealed that 566,700 more crimes had been committed in 1970 than in 1969.

Admittedly confused, I sought out my friend Professor Heinrich Applebaum, the great Justice Department statistician, whose definitive book *Do Decimal Points Have a Sex Life?* is used in every math class in the country.

"Professor Applebaum, the Justice Department reports that the rate of crime has gone down in the country under President Nixon. Yet the same report says there has been a million more crimes in the last two years. How can that be?"

"It's quite simple," said Applebaum. "Percentagewise crime has gone down, crimewise it's gone up."

"But where does that leave the average person?"

"It depends whether you're a Republican or a Democrat. If you're a Republican you have nothing to fear walking the streets of our American cities. But if I were a Democrat I'd stay home."

"Are you saying that the Republicans are trying to take the crime issue out of the 1972 campaign?"

"They have," Applebaum said. "The last year the Democrats were in office crime had gone up 13.8 percent. When the Republicans took over in 1969 it only went up 12.0 and last year 11.3 percent. The Democrats can't argue with that."

"But still more people were robbed, mugged, murdered, and raped in 1969 and 1970 than they were in the previous years."

"We're not talking about people," Applebaum said, irritated. "We're talking about percentages. You can't think about the people who were molested in 1969 and 1970; you have to think about the ones who weren't mugged . . . this year thanks to President Nixon's leadership."

"It's hard to think in those terms," I admitted.

"That's because you're not running for election next year. You must understand the reporting of crimes is a very serious business and can cause great conflict. J. Edgar Hover, in order to prove he is doing his job, has to show that crime is going up in the country. At the same time the administration has to prove that crime is going down."

"The Attorney General has solved the problem by reporting the percentages, which are lower, and Hoover, by reporting the crimes, which are higher. That's the beauty of statistics. It makes everyone feel better."[1]

---

[1] Reprinted with permission.

The preceding levity serves a useful purpose in addition to the provision of humor. Statistical information is sometimes used as a prop for conclusions unwarranted by the data. After studying this text, the reader should have an improved awareness and ability to detect improper use of statistical methods.

Inferential Statistics

Inferential statistics is a body of techniques used in drawing conclusions concerning some group on the basis of incomplete knowledge. The group of interest is usually termed the population or universe:

> *population:* the set of elements on which information is desired.

The population under study can be relatively small (e.g., the 250 salaried employees of the Normal Company) or very large (e.g., the 70,000 industrial customers served by Midstate Power Company). Of course, the population in a particular study should be clearly identified, or the aims of the investigation may not be realized. Sometimes the desired information on a population is obtained by means of a census.

> *census:* the individual observation of the entire population.

One example of a census is the decennial Census of Population and Housing directed by the U.S. Bureau of the Census, but note that the definition above would encompass much more than federal efforts. Measures computed on the basis of a census are termed *parameters* and are usually symbolized by Greek letters.

Pragmatism frequently dictates that information concerning a population be obtained in a manner other than the census. Time and monetary considerations are important factors in such a decision. Or occasionally one may encounter measurements that can be obtained only through destructive testing; for example, the life of a light bulb can only be determined by recording the time until burn-out. In such a case, a census is obviously out of the question. Hence, information or conclusions concerning some populations will be based on incomplete knowledge (i.e., something less than census information). For the statistician, this partial knowledge will typically be obtained via sampling.

> *sample:* that portion of the population available for analysis.

A measure computed on the basis of sample data is termed a *statistic* (note the singular) and usually symbolized by a Roman letter. A major portion of this text is concerned with sampling and the use of sample data to reach conclusions concerning the population.