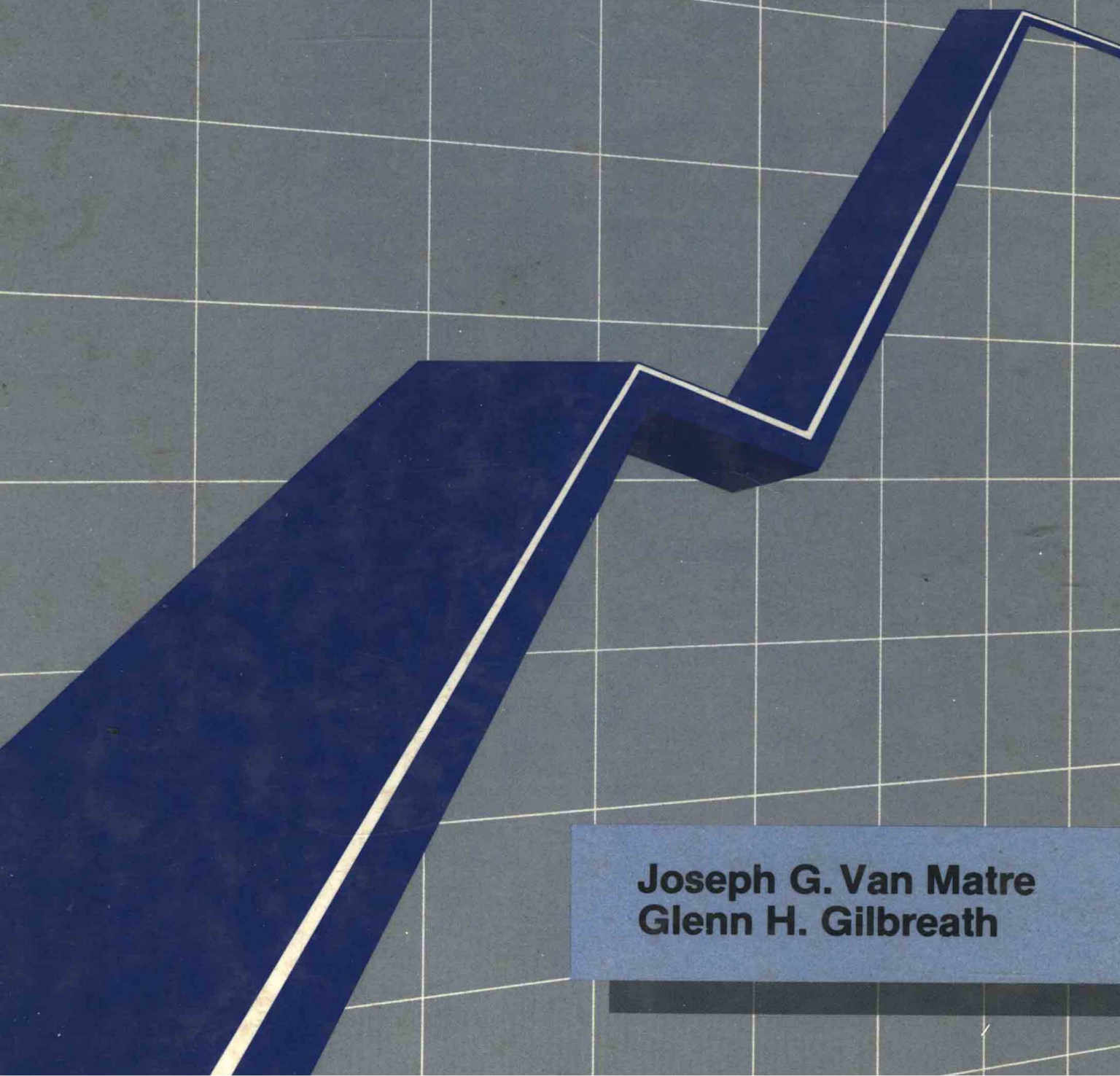
Statistics for Business and Economics

Third Edition



Joseph G. Van Matre
Glenn H. Gilbreath


BPI
IRWIN

ISBN 0-256-03719-1

Library of Congress Catalog Card No. 86–72249

Printed in the United States of America

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

Preface

Statistics for Business and Economics, third edition, is designed primarily for the introductory statistics course in business, economics, and management. Our goal is to provide readers with both an understanding of and appreciation for statistical concepts. Our method for achieving this goal is to present both *concepts* and *applications*. Illustrations of the concepts are plentiful. More comprehensive "applications" (often called case studies) are also provided throughout the book. Reprints of relevant material from business periodicals are included. A generous number and variety of exercises are provided at the end of each chapter.

The use of computers and statistical software packages is integrated throughout the text as a natural component of statistical analysis. Widely available statistical software such as Minitab, SAS, and SPSS are used for illustration. We also show how the popular electronic spreadsheets such as Lotus 1-2-3 can assist in organizing and analyzing data. Nevertheless, the text can be used in its entirety without access to computing equipment. Readers that skip the computing exercises can still benefit from the illustrations provided and will have some familiarity with computer-generated output. While widely available computer routines provide efficient ways to perform the more laborious calculations, the use of a computer routine is not a substitute for sound statistical reasoning; it is merely a computational aid.

To the Student

We have attempted to make your introduction to statistical methods both interesting and informative. Modern managers must use the results of statistical analysis in making business decisions. This text is not designed to make you a statistician. Instead it is designed to provide sufficient familiarity with statistical methods to make you a knowledgeable user of statistical results.

The method of presentation is to first introduce concepts and follow these by illustrations. Insight into more in-depth use of certain concepts is provided in the "applications" distributed throughout the text. Reprints of related discussions from various business sources provide additional insight into current practice. Historical consciousness provides a well-rounded perspective for any subject. To provide

this, we offer biographical sketches of some important contributors to the development of statistical thought.

We suggest that you study each section by carefully reading the text and working through the illustrations presented. Each chapter concludes with a summary of major topics. Convenient lists of important terms, symbols, and formulas are provided for review and reference. The true-false and fill-in-the-blank questions may be examined to reinforce major points, but the best learning tool is devoting as much time as possible to working the exercises at the end of each chapter. You will find answers to the odd-numbered exercises at the end of the book.

An optional Study Guide/Workbook to accompany the text is available. This provides a concise review of each chapter along with additional illustrations, questions, and exercises. Those who desire additional reading on the topics presented in the text may consult the references listed on the last page of each chapter.

To the Instructor

The introductory statistics course provides the sole exposure to statistics for many students, while others may continue study in this and related areas such as management science and econometrics. To meet the needs of both groups, this book emphasizes a conceptual approach with both verbal and mathematical explanations. Selected derivations are presented, usually in footnotes or in the appendixes. Illustrations and exercises are plentiful; the amount of "busy" work is kept to a minimum.

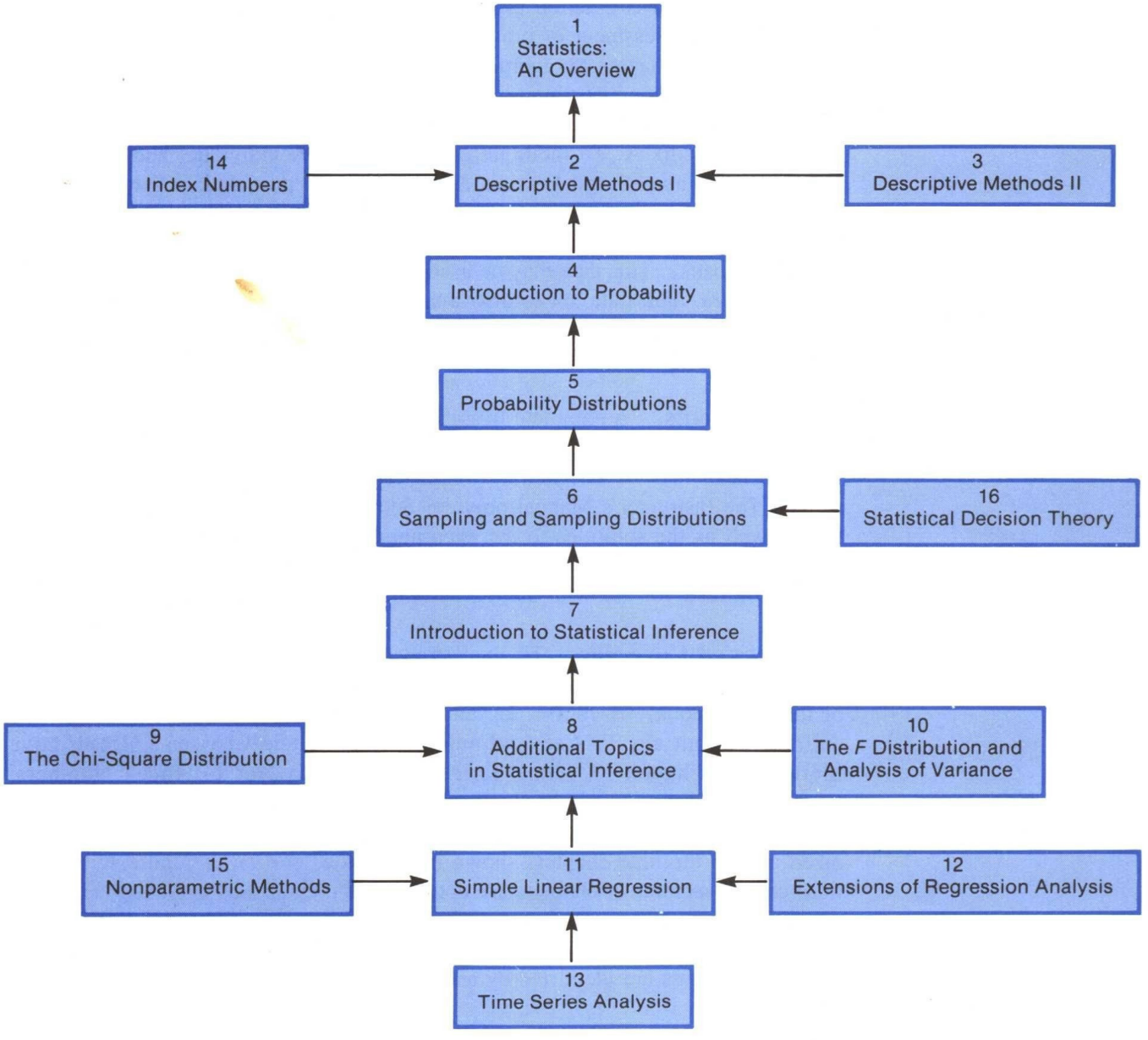
We assume the reader's level of mathematical maturity is college algebra. Students with a knowledge of elementary calculus will benefit from the supplementary explanations given for certain concepts. However, calculus-based material is relegated to footnotes or optional sections so that no loss of continuity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

Overview of the Book. All elementary topics usually considered under the general heading of "descriptive statistics" are presented in Chapter 2, Descriptive Methods I, including frequency distributions, averages, and measures of variation for a single variable. Chapter 3, Descriptive Methods II, is a completely new chapter, devoted primarily to bivariate data and including such topics as crosstabulation, sorting, graphics, and descriptive measures such as covariance and correlation.

Probabilistic concepts are essential for an understanding of statistical inference. A careful presentation of the basic concepts is given in Chapter 4, while probability distributions 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-

FIGURE 1: CHAPTERS IN BOOK WITH ARROW POINTING TOWARD PREREQUISITES



square, and analysis of variance. Chapter 16 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 13, includes exponential smoothing and a discussion of quadratic and exponential trends.

A comprehensive set of tables, more extensive than in previous editions, is given in Appendix A. To facilitate ease of use, both individual and cumulative terms of the binomial and Poisson distributions are provided; and the binomial tables include values of $\pi > 0.5$. In addition, another appendix includes the *Business Week* "Bank Scoreboard" that provides data on the characteristics of the largest 200 banks. This data may be used in a variety of ways for both illustrations and exercises. The data set is also available in diskette form for adopters by contacting Business Publications, Inc., 1700 Alma Road, Suite 390, Plano, TX 75075.

Supplements. A complete Instructor's Manual and an optional Study Guide/Workbook are available. The guide may be used independently by students as a supplement or may be 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 even-numbered exercises in the Study Guide. Transparency masters for selected text material are included in the manual. A Manual of Tests for use with the text is also available to adopters.

Suggested Coverage. The book may be used for courses of varying length. For a course over two academic terms, we recommend a fairly complete coverage of the entire book. Of course, certain chapters are optional and can be omitted entirely without any loss of continuity. These include Chapter 3, Descriptive Methods II; Chapter 13, Time Series Analysis; Chapter 14, Index Numbers; Chapter 15, Nonparametric Methods; and Chapter 16, Statistical Decision Theory. A one-term course covering descriptive and inferential statistics could be organized around the nine chapters shown in the center portion of Figure 1. In addition, certain sections of chapters are optional and may also be excluded. Then, for example, one could treat probability rather lightly and gain time to spend on bivariate description or chi-square. Optional sections within chapters as well as optional exercises are identified by asterisks.

Acknowledgments

To all those individuals who took the time to send comments and suggestions about previous editions, we sincerely appreciate your thoughtfulness and efforts. Be assured that all comments were welcomed and carefully studied. Without doubt, this helped to improve the content. We also benefited from the detailed and thoughtful reviews of Ron Barnes, University of Houston; James Bartos, Ohio State University; Ronald S. Koot, Pennsylvania State University; Terry Ridgeway and William Robinson, University of Nevada, Las Vegas; and Mary Whiteside,

University of Texas at Arlington. For assistance with PC statistical software, we thank Jeff Ballowe and Jeff Wiss, SPSS; Eve Goldman, SYSTAT; and Vicki L. Jodon, Minitab. We very much appreciate the helpful suggestions and support of all our colleagues at the University of Alabama in Birmingham and Virginia Commonwealth University. In particular, thanks to U.A.B. colleagues M. Gene Newport, David Lewis, and Keith Bryant for help in finding a quiet place to work; Ted Bos and Seung-Dong Lee for help with the PC hardware; and Ted also for resolving software problems and critiquing some of the new material. Jennie Alley, U.A.B. graduate assistant, has been very helpful in many phases of this work. At V.C.U., we particularly wish to thank Robert L. Andrews, Don M. Miller, Paul M. Umberger, and graduate students Ash Ghouse and Stuart Parks.

For typing and clerical support, we are very grateful to Dot Haufler and Jill Kramer, both of Virginia Commonwealth University, and to Rose McCoy and Betty Smith, University of Alabama in Birmingham.

We especially thank T. Hillman Willis, Louisiana State University in Shreveport, for preparing the Workbook/Study Guide; and Dwayne Key and Lynette Solomon, Stephen F. Austin State University, for preparing the Manual of Tests.

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.

Finally, we wish to thank our family members, Mindy, Kelly, Dru, and Catherine, for their patience, understanding, and support during the many hours that we were unavailable.

Joseph G. Van Matre
Glenn H. Gilbreath

Contents

**Optional material.

**Optional material.

**Optional material.

**Optional material.

'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 you begin this study of statistics, you must be apprised that many new concepts lie ahead. You 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, you will find a variety of new symbols (e.g., σ, μ) and words (e.g., median, regression). The Greek alphabet printed in Appendix B 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 you complete this book, your statistical knowledge will surpass Alice's understanding of longitude and latitude.

1.2 STATISTICS DEFINED

The word **statistics** means different things to different people. For example, statistics are Marcus Allen's yards per carry and season touchdowns, or the price-earnings ratio of a stock. However, as you shall see, statistics is much more than merely 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 this is a rather broad definition, it is useful to 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 you were given the SAT scores of the 300 freshman business students and 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 classic 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 there 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 Hoover, 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.

Amid the humor of this column is a serious message. Statistical information is sometimes used as a prop for conclusions unwarranted by the data. After studying this text, you should have the awareness and ability to detect an improper use of statistical methods.

Statistical Inference

Statistical inference 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). Sometimes the information desired on a population is obtained by 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 (e.g., σ, π).

Pragmatism frequently dictates that information concerning a population be obtained in a manner other than a 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 burnout. 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). This partial knowledge is typically obtained through 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 or italic letter (e.g., s, p). A major portion of this text is concerned with sampling and the use of sample data to reach conclusions concerning the population.