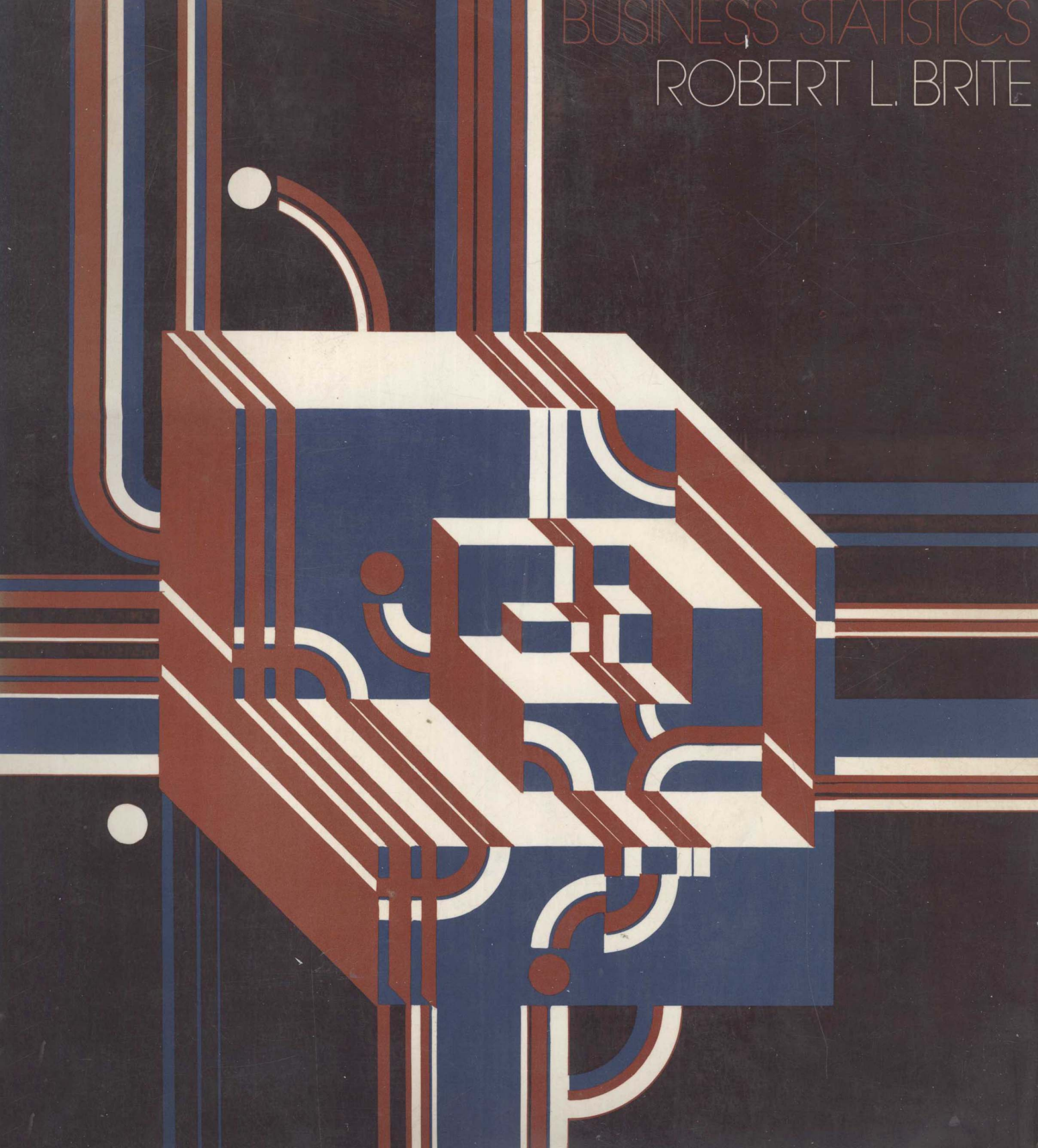
BUSINESS STATISTICS

ROBERT L. BRITE



Business Statistics

Robert L. Brite
University of New Orleans

ADDISON-WESLEY
PUBLISHING COMPANY
Reading, Massachusetts
Menlo Park, California
London
Amsterdam
Don Mills, Ontario
Sydney

Library of Congress Cataloging in Publication Data

Brite, Robert L.
Business statistics.

Includes index.
1. Commercial statistics. 2. Statistics.
I. Title.
HF1017.B68 658.4'033 79-25657
ISBN 0-201-00561-1

ISBN 0-201-00561-1
ABCDEFGHIJ-MA-89876543210

Preface

This text presents a clear and concise treatment of the basic concepts of business statistics. It is written in a manner that will appeal to the beginning student of statistics. Treatment of both introductory and complex topics is logical and intuitive rather than mathematical. By minimizing the mathematical background necessary, the emphasis is on student understanding of concepts rather than the learning of mathematical proofs. Although the text is written at a basic level, its emphasis is on decision making in business. It leads the student from the simple concepts of computing the mean and median through statistical techniques applicable to decision making.

This book is intended for use as a two-semester text at the sophomore level. However, the text is written in such a manner that a one-semester course may have great flexibility in including latter chapters, since most are independent of earlier material. The material in the text is self-contained, requiring only basic algebra as a background.

The book focuses on the application of statistics to decision making in the business world. Many examples from the business world are included to realistically illustrate the ways in which statistical methods are employed. Students completing this course will not be statisticians, but they will have an appreciation and firm understanding of basic statistical concepts.

The work-text approach is used in the text, and you will find that it encourages interest and independent mastery of the material. Behavioral objectives listed at the beginning of every major section within each chapter alert the student to what is most significant in the discussion to follow.

These objectives also serve as a handy summary and guide for review. Marginal exercises appearing "on location" give students the opportunity to work problems illustrating the very concepts just introduced in the text. The work text should not be confused with the "workbook." The work text contains all text material plus supporting marginal exercises. The work-test approach has been found effective in mathematics instruction and is easily adaptable to the teaching of statistics. Complete answers to all marginal exercises appear at the end of the text. Students can check their answers, and if they have not worked a problem correctly, they will be able to determine where they made their mistake. Answers are also included for the odd-numbered end-of-chapter problems.

The text is organized into three basic areas: descriptive statistics, covered in Chapters 1–5; probability and statistical inference, covered in Chapters 6–13; and regression analysis and its applications, covered in Chapters 14–15. Chapters 16–18 treat chi-square, analysis of variance, and nonparametric statistics in an intuitive and nonmathematical manner. Chapter 19 covers the basic components of time-series analysis, and Chapter 20 includes a logical treatment of decision theory. The section on descriptive statistics includes methods of data presentation and computation of measures of central location and variation. In the second section, probability is presented in an understandable, nonmathematical manner intended to emphasize the applications of probability rather than abstract concepts. Bayes's theorem is also included. Through use of the intuitive approach, sampling theory is treated in such a way as to minimize mathematical derivation. An intuitive approach is also used to explain why it is important to understand sampling distributions. Estimation and hypothesis testing are presented with many examples illustrating the use of these statistical tools in the decision-making process. In the third section, simple regression analysis is covered. Emphasis falls on understanding the *concept* of regression and correlation analysis, as well as the computational procedure necessary. A step-by-step approach helps the student learn the computational methods and understand their significance. Treatment of the time-series and simple regression analysis is included as well. A brief treatment of multiple regression emphasizes interpretation of the multiple regression coefficients and use of the computer in computing multiple regression coefficients. Chi-square and analysis of variance are treated by using the step-by-step approach, and basic nonparametric tests are discussed. The step-by-step method is also used in the time-series and decision-theory sections.

The primary aim of this text is to present statistics in a more interesting and less traditional manner. If students become involved with the material (as the work-text format requires), they will find their study of this text an enjoyable learning experience. Many students face statistics with some anxiety and displeasure. The work-text format amply illustrates the use of essential techniques and makes computations easier to learn. And it

ensures that all major theories and concepts are mastered in the context of their concrete applications in the business world, where the student will use them. In later courses, students will be exposed to other techniques that build on the foundation of this course. This text has been written in the hope that it will help students comprehend and master difficult material. The goal has been to make the material as easy to understand and remember as possible.

I am 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 tables from their book *Statistical Tables for Biological, Agricultural, and Medical Research* (6th edition, 1974).

New Orleans, Louisiana R.L.B.
January 1980

Contents

Introduction 1

SECTION 1.1 WHAT IS STATISTICS?

The term "statistics" may be used in two different ways. You are probably already familiar with one of the meanings. *Statistics* are numbers that have been collected and are available for our investigation. For example, we are familiar with baseball batting averages, yardage gains of leading professional football players, and salaries earned by top musical entertainers. These numbers or data are referred to as statistics. They have been collected for a specific purpose.

Statistics also has a much broader meaning encompassing a field of study that involves the collection, analysis, and presentation of data in business and other fields. It also involves the interpretation of such data by the user in the business world in order that information may be used in making decisions. Often, businesspersons make decisions without complete information. The study of statistics is intended to cut down on such practices and provide a sound base for decision making. When you have completed this course, you will know *what* information is important to the decision maker, *how* to collect the information, and *how* to interpret the information accurately.

■ **Definition: Statistics** | ***Statistics is a field of study involving the collection and analysis of data for use in making decisions.***

Statistics is normally divided into two separate areas. First, we have *descriptive statistics*, which involves the collection and manipulation of data. This includes such things as computing averages, drawing graphs, presenting tables, etc. Many people think that this is all there is to statistics. However, descriptive statistics is perhaps of lesser importance than the area known as statistical inference. While we must use descriptive statistics in order to be able to make decisions, we find that the decision-making aspect of statistics is much more important.

■ **Definition: Descriptive Statistics** | ***Descriptive statistics is that part of statistics involved with the collection and manipulation of data.***

The second area of statistics is *statistical inference.* Statistical inference is that part of statistics which deals with drawing conclusions based on the statistical information gathered. The decision-making process will usually involve uncertainty. Uncertainty may be included as a part of our decision model through the use of probability. Thus we have two basic areas of interest—descriptive statistics and statistical inference—but we also must consider *probability,* which links the two.

■ **Definition: Statistical Inference** | ***Statistical inference is that part of statistics which involves the drawing of conclusions about the population based upon a sample.***

Normally, data collection does not involve a census. If a census technique is used, then all members of the population must be included. This is not only extremely time-consuming and expensive, but also unnecessary in most cases. A sample will normally be selected from the population in order to gain the needed information. The *population* is defined as the entire group in which we are interested. A *sample* is a subset of that population. While a sample may be chosen in any number of ways, we would usually expect a representative sample to be the most useful. The sample that best represents the population will be more useful in decision making than one that does not. The best method of obtaining a representative sample is to take a random sample. A *random sample* is a sample chosen in such a way that every member of the population has an equal chance of being chosen.

■ **Definition: Population** | ***The population is the entire group about which the statistician is interested in drawing a conclusion.***

■ **Definition: Sample** | ***The sample is a subset of the population. It is chosen as a representative of the population.***

■ **Definition: Random Sample** | ***A random sample is one chosen in such a manner that every member of the population has an equal chance of being included in the sample.***

▶ **Example 1.1** We are interested in studying the spending habits of college freshmen. Thus our population can be defined as all college freshmen at the various universities within the country. A number of different types of samples may be taken from this population, but a random sample probably would be the most representative. A random sample must be taken in such a way that each member of the population has a equal chance of being chosen. All must be included as potential members of our sample. How will we do this? We must either collect the names of (or somehow identify) all freshmen in the country. We then place all of these names in a bowl and draw, or we have a computer select the sample. Normally, we will find that a sample of a few hundred will be adequate to represent a very large population. Many times smaller samples will be adequate, depending upon the type of information being collected. ◀

It is important to be able to distinguish between a sample and a population. Let us assume that we have a population shown by the diagram in Fig. 1.1.

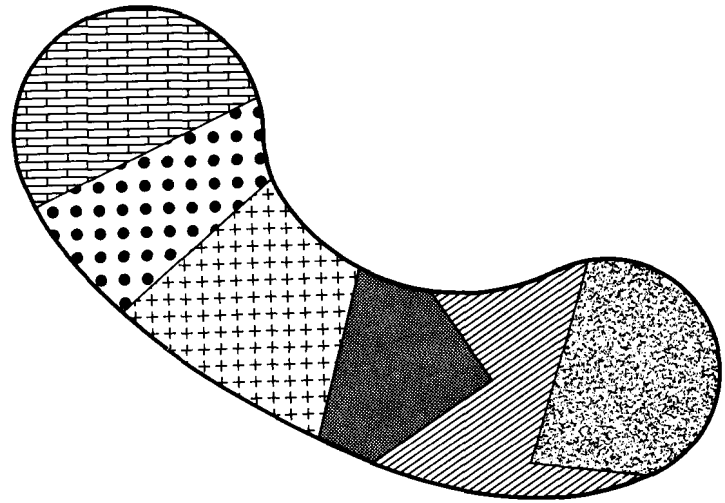


Figure 1.1 Population

A number of possible samples could be selected from this population. But, we notice that different parts of the population are of different texture. In order to gain an adequate representation of the population, we certainly would not want to choose our entire sample from one texture. We wish to select, by choosing randomly, with each group having an equal chance of being chosen, a sample that might look like Fig. 1.2.

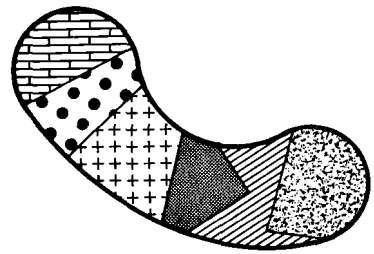


Figure 1.2 Sample

Note that the proportions of the textures in Figs. 1.1 and 1.2 may not be precisely the same. This usually will be true in sampling. No matter how hard we try, it usually is not possible to get perfectly representative samples from the population. However, as long as our sample data are not continually biased in the same direction, thus favoring certain population characteristics, we view the sample as at least an adequate and usually an excellent

representative of the population. The difference between the sample and the population is referred to as *sampling error,* which will be discussed in detail later.

■ **Definition: Sampling Error** | ***Sampling error is the error in the sample which occurs because of sampling. It accounts for the differences between the sample and the population when a random sample is taken.***

The population is the body about which we are interested in making an inference. We usually will work from the sample information in order to make such an inference. We must be careful, however, to be sure that we identify our population correctly and that we do not define the population as broader than the group from which the sample is collected. For example, if we collected a sample from the south only, we should not make an inference about the entire nation. Such incorrect inferences are often made. To ensure that this will not happen, the population to which the sample applies must be defined accurately.

Some populations are *infinite* (unlimited in size) and some are *finite* (limited in size). The population of individuals in the world or even in the United States (although actually finite) is, for all practical purposes, infinite. It is so large that we are not likely to exhaust the population or even to come near obtaining a sample of significant size relative to the size of the population. On the other hand, the number of business firms in a given city is finite. The number of firms definitely is limited and a census is possible. When sampling from a limited or finite population, we must make some adjustments. This will be discussed in later sections.

Although in this section you have not learned exactly what statistics is, we hope you have gained some appreciation for a number of the topics that we will be discussing. In succeeding sections of this chapter, you will be introduced to some of the basic uses of statistics and given some information concerning what the course is all about, as well as how you should study statistical material.

SECTION 1.2 HOW IS STATISTICS USED?

There are many applications of statistics in the business world. The most common usage is the collection of information concerning the operation of a business firm in order to analyze its position and make decisions concerning its future. Descriptive statistics is particularly important in the computation of relevant "statistics" to compare our present situation to that of the past and, possibly, to that of our competitors. We will find that the decision-making portion of the text, involving hypothesis testing and estimation, is very useful in formulating a concept of statistical significance

and making a meaningful decision. Regression and correlation analysis will be used to analyze relationships between two or more variables. We will see that by being able to formulate such a relationship, we may be able to forecast future events based on a relationship between explanatory variables or a time trend. Thus statistics is very important to the decision-making process. In order to collect accurate, useful information, the decision maker must have the correct background information.

SECTION 1.3 THE STATISTICS COURSE

How to Study Statistics

Statistics is a course not unlike mathematics. It differs from math in that there is more intuition involved. However, both areas require a great deal of problem-working and self-teaching. Few students can master the material in a statistics course without working a multitude of problems. Over the many semesters in which I have taught this course, I have developed a success pattern for students to follow in learning this material. If you follow this plan, it is difficult not to succeed in the course.

First, you must come to the course adequately prepared. A minimum amount of mathematical background is necessary. However, if you do not understand the material in Chapter 2, then additional preparation is necessary, either through review or through a formal course.

Second, you must commit yourself to mastering the material. Many students find that statistics is boring. This is true if you remain totally uninvolved. However, once you consider the number of applications available from this material, the course can be interesting and exciting. You will find that few courses will be of more use to you in the years to come.

Third, you must be willing to read the text material and work the exercises as assigned. *While you may feel that you understand the material based on the teacher's explanation, no amount of "watching the teacher do it" will substitute for doing it yourself. Be sure that you, on your own, can work every problem assigned.* If you do not understand how to work a problem, refer to textual material or question the teacher. You will find the problems at the end of each chapter very important in determining your mastery of the material. This can not be overemphasized. You *must* work the problems in order to learn the material. The problems are designed to illustrate some fairly difficult concepts. Working the exercises will make learning the concepts much easier.

Fourth, go to class prepared. Study the material and attempt to work the problems prior to going to class. Determine the areas in which you have questions. Clear up the misunderstanding in the class. You must learn the details on your own. Once you have attended class and cleared up any questions, then rework the exercises that you did not understand.

Fifth, take the self-tests at the end of the text conscientiously. Do not "cheat" on these. While working the exercises earlier in the chapter, use the text freely. However, in working the tests, pretend you are in an actual test situation. Grade yourself on the tests and then rework those problems that you did incorrectly.

You will find that it is essential that you keep up with the material. This is a course in which it is not possible to "cram" for the exam. Since each class will build on the previous class, you will find that the cramming procedure will mean that your classes are largely worthless because you did not have the adequate preparation to attend the class. Set up a program and follow it. Plan your own quizzes; give yourself deadlines and stick to them.

What is Included in the Course?

In the succeeding chapters, you will work through a great deal of material. The first section of material is descriptive in nature and involves the basic mathematical background needed for the course. This section is optional, depending upon your mathematical background and abilities. You then will learn how data are collected and presented. You will see that there are a number of summary statistics that we may use to present data more meaningfully. We then move into the area of probability. There are a number of chapters dealing with probability and probability concepts. Some of these concepts are difficult. You probably will find this section the most difficult to master. It is important, however, because it links the material that you just learned concerning descriptive statistics with the later materials concerning decision making. Mastering probability requires logical thought. You must be able to set up a problem and proceed logically through the problem in order to learn the material. There is also a section on sampling. We will see that sampling is essential to the decision-making process and is used in situations where we could not possibly take a census. We will see that sampling distribution has particular characteristics that greatly facilitate our use of most economic and business data.

All of this material which you will have learned leads up to the final chapters. Chapters 10 to 20 are the really important chapters in the course. They involve statistical inference and decision making, the collection of data, the presentation of data, the measuring of summary statistics, and probability. All of this leads us to the ability to make decisions under conditions of uncertainty. In these chapters, we will see how to estimate characteristics of the population based on the sample information, and how to use sample information to test hypotheses concerning the population. Next, we will cover linear regression and correlation analysis. This material involves the analysis of paired data, where two variables are involved and we wish to know the association or functional relationship between the variables. The final topics are chi-square, analysis of variance, and non-