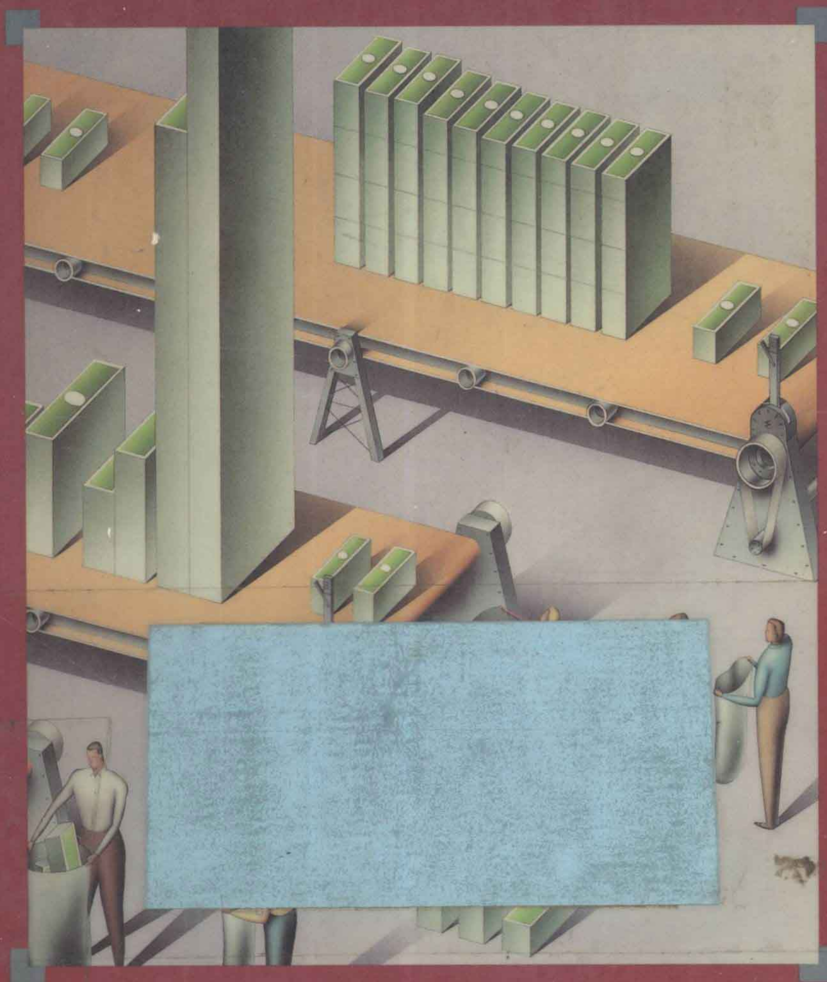
BASIC BUSINESS STATISTICS



E. J. FREED

BASIC BUSINESS STATISTICS

E. J. FREED
University of Portland

Harcourt Brace Jovanovich, Publishers
San Diego New York Chicago Austin
Washington, D.C. London Sydney
Tokyo Toronto

ISBN: 0-15-504900-3
Library of Congress Catalog Card Number: 90-82260
Printed in the United States of America

To Sally, Amy, and Dan

PREFACE

Basic Business Statistics is intended to provide the basis for a single-semester first course in statistics. It can also serve as an easy-to-read review for students who might want to reacquaint themselves with elementary statistics in preparation for a class requiring the application of basic statistical principles. The informal style and commonsense explanations that characterize this textbook should help to reduce the anxiety that frequently marks a student's first exposure to the field. The intent here is to provide a book that supports rather than threatens, one that will offer students who possess modest-to-good mathematical skills a resource that is interesting, intelligible, and perhaps even a little entertaining.

Unlike many competing books, which are often no more than abridged versions of textbooks intended for use in a two-semester course, *Basic Business Statistics* was conceived and designed specifically for a one-semester course in business statistics. The emphasis is on simple illustrations rather than on formal proofs and theorems. Potentially confounding mathematical notation is kept to a minimum. And a serious effort has been made to help students distinguish between core issues and those of secondary importance, something absolutely critical in promoting student confidence and comprehension. Particular attention has been paid to the development of a clear, comprehensible introduction to *confidence interval estimation* and *statistical hypothesis testing*—two crucial areas of study in beginning statistics. The nature and role of sampling distributions are carefully established and continually reinforced.

Several features have been incorporated into *Basic Business Statistics* that are specifically designed to aid students in their attempts to master statistical concepts and procedures.

- *In-Chapter Exercises* are included in each chapter to provide students with an immediate opportunity to apply the concepts covered in the chapter.
- *Problem Sets* have been designed to augment and extend chapter discussions, rather than simply to provide opportunities for drill and repetition. The end-of-chapter exercises cover a broad range of applications that are both challenging and interesting.

- *Chapter Supplement Sections* are provided for the chapters covering Descriptive Statistics (Chapter 2), Probability Theory (Chapter 3), Statistical Inference (Chapter 5), and Regression Analysis (Chapter 9). These optional sections are included for instructors who wish to cover these topics in greater depth.
- *Comprehensive Solutions*—not just answers—are provided in the book for nearly half of the end-of-chapter problems and for all the in-chapter exercises.
- *Chapter Objectives* give students a clear understanding of what they can expect to encounter in the material that follows.
- *Chapter Summaries* reinforce and clarify the chapter material that has just been covered.
- *Key Terms* for discussion and retention are listed at the end of the chapter in which they first appear.

Basic Business Statistics is not intended to be all things to all people. What it does offer is an effective alternative to the sometimes-imposing tomes that are poorly suited to a one-semester format. Extensive class testing of the manuscript at the University of Portland has been extremely useful in the attempt to achieve this and other ends.

Ancillaries

A *Solutions Manual* containing completely worked-out solutions to all in-text exercises and end-of-chapter problems is available.

A *Testbook* containing true/false, multiple-choice, and other problems is available.

Acknowledgments

Professor William Maurer (Naval Postgraduate School) and Professor Richard Spinetto (University of Colorado), by their example, provided me with a model for effective teaching. Professor George Chou (University of Portland) provided support and encouragement throughout the writing of the book. Professor James Jurinski (University of Portland) was helpful in finding a suitable publisher. And Theresa Vu, Linnea Aguiar, and Dan Freed provided valuable assistance in preparing problem solutions and in checking computational accuracy. My thanks to all of them.

I also thank Mickey Cox and Scott Isenberg of Harcourt Brace Jovanovich for their efforts in transforming a rough outline into a finished product. Finally, I thank the reviewers for their many constructive comments and criticisms: Dan Albrechtson, Milwaukee Area Technical College; Les Dlaby, Lake Forest College; Darrell Christie, University of Wisconsin–Stevens Point; Gerald Evans, University of Montana; W. F. Mackara, East Tennessee State University; and Norman Rittgers, Pasadena City College.

CONTENTS

CHAPTER 1

AN INTRODUCTION TO STATISTICS

CHAPTER OBJECTIVES

Chapter 1 should enable you to:

1. Define the term *statistics.*
2. Identify the two main branches of statistics: descriptive statistics and inferential statistics.
3. Appreciate the role of probability theory in statistical analysis.
4. Understand the nature of data and the concept of levels of measurement.
5. Anticipate the topics to be discussed later in the text.

1.1 STATISTICS DEFINED

In common usage, the term *statistics* can take on a variety of meanings. For example, it's frequently used to describe numerical data of almost any sort—height, weight, batting average, GPA, age, flash point, and the like. Some people might connect the term to the results of surveys, polls, and questionnaires. We'll use *statistics* in this text primarily to designate a specific academic discipline focused on methods of data collection, analysis, and presentation. In nearly every case,

> **Statistics** involves the transformation of *data* into *information.*

That is, statistical analysis—in whatever form—generally entails translating raw numbers into meaningful information in order to establish the basis for sound decision making.

1.2 BRANCHES OF STATISTICS

It's possible to divide the discipline of statistics into two principal branches: descriptive statistics and inferential statistics.

Descriptive Statistics

Descriptive statistics focuses directly on the summarization and presentation of data. Here, we're typically looking for effective ways to refine, distill, and describe data so that, as we've discussed above, the numbers involved become more than simply numbers. While descriptive statistics can potentially involve some rather complex measures, we will, at the introductory level, deal in measures not much more complicated than simple averages.

Inferential Statistics

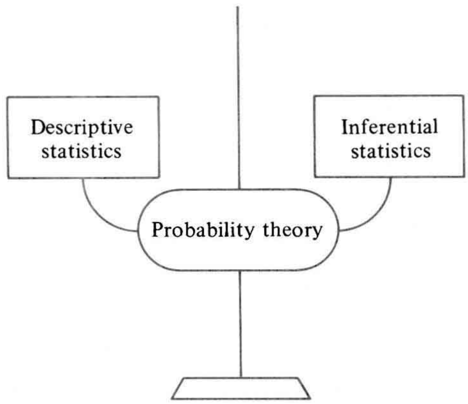
When dealing with especially large data sets, it's often necessary, for reasons of time, cost, or convenience, to draw conclusions (or to make *inferences*) about an entire data set by using only a subset of the values involved. **Inferential statistics** (sometimes called **inference statistics** or **inductive statistics**) deals with the selection and use of *sample* data to produce information about the larger *population* from which the sample was selected.

Note: The term **statistic** is sometimes used to refer specifically to characteristics of a sample. For example, the average value in a sample of 10 measurements could be considered a sample "statistic." In contrast, population characteristics are frequently referred to as **parameters**. In inferential statistics, we'll routinely use sampling theory to relate sample "statistics" to population "parameters."

1.3 PROBABILITY

Probability theory—the study of likelihood or chance—can be seen as the link between descriptive and inferential statistics. A knowledge of basic probability will enable us to connect what we see in a sample to what we would *likely* see in the population being represented by the sample. Figure 1.1 details this simple classification scheme.

FIGURE 1.1 **BRANCHES OF STATISTICS**



1.4 A BROADER PERSPECTIVE

To put statistical analysis into a somewhat broader context, we can identify two distinct analytic views of the world—two sets of beliefs about "how things are." In a *deterministic* view of the world, the world we see is essentially a world of certainty. It's a world that consists primarily (if not exclusively) of one-to-one, cause-and-effect relationships: given some cause, there's a unique and identifiable effect; given some effect, there's a clearly discernible cause. The world here, while not necessarily simple, is at least well defined.

In contrast, under a *probabilistic* view of the world, things don't appear quite as straightforward. Here, given some cause, there may well be a number of possible effects; given some effect, there may be a number of possible causes. In this conception of reality, the world is dominated by a real, and complicating, *un*certainty.

It turns out that for each of these two contrasting world views, an appropriate set of analytic tools is available. The tools of probability and statistics (especially inferential statistics) are designed to deal with problems arising in a distinctly uncertain world. (We can, in fact, think of probability theory as the means to measure—and thus control for—uncertainty.) In contrast, optimization tools such as linear programming and differential calculus are better suited to deal with problems encountered in a "certain" (i.e., deterministic) world.

1.5 THE STRUCTURE OF THE TEXT

The chapters that follow are intended to introduce the student or practitioner to the principles of basic statistical analysis without the rigid formality of many of the books written in the area. In general, we've tried to use instinctive, intuitive arguments to replace formal proofs and theorems. For the most part, key concepts and procedures are developed around simple illustrations. Excessive mathematical notation is avoided. At the same time, we've tried not to exclude complex issues from the discussion.

Importantly, substantial effort has been made to separate "wheat" from "chaff"—key and fundamental issues versus issues of only secondary or supporting significance. (In my own view, failure to differentiate clearly for the beginning student those ideas that are absolutely fundamental from those that are peripheral to the central theme is the main deficiency of most introductory texts in the field.) We've attempted to keep each chapter brief and to the point. Problem sets have been designed to augment and extend chapter discussions rather than solely to provide opportunities for drill and repetition. Solutions to nearly half of the end-of-chapter exercises are provided in the back of the text.

For a class proceeding at a moderate pace, the material presented fits readily into a single-semester, introductory-level course. What follows is a capsule summary of the remaining chapters.

Chapter 2: Descriptive Statistics

Chapter 2 introduces basic descriptive measures useful in converting data into information. Included are measures of center (or central tendency)—the mean, the median, and the mode, as well as measures of dispersion—the range, the mean absolute deviation, the variance, and the standard deviation. The chapter supplement adds a measure of relative dispersion (the coefficient of variation) and two measures of position—percentiles and quartiles. Data presentation possibilities are also described, with a primary focus on simple bar charts (histograms).

Chapter 3: Probability Theory

Chapter 3 outlines a number of basic probability concepts: simple, joint, and conditional probabilities; statistically independent and mutually exclusive events;

and the two principal rules of basic probability—the additive and the multiplicative rules. Visual formats—probability trees, Venn diagrams, and joint probability tables—are shown to be useful in organizing and solving a variety of problems.

Chapter 4: Probability Distributions

Probability distributions, which involve the comprehensive listing of probabilities assigned to a full range of possible events, are viewed as a logical extension of Chapter 3 principles. Special-purpose distributions (the binomial, the normal, and the Poisson), which can be shown as compact mathematical functions, are introduced in order to efficiently produce probabilities under certain clearly defined conditions. The chapter supplement adds the exponential distribution to the special-purpose list.

Chapter 5: Statistical Inference: Estimating a Population Mean

Chapter 5 initiates a discussion of basic sampling procedures. Here statistical theory is introduced in order to link sample results to the population from which the sample was selected. The focus is on the construction of **confidence interval estimates** of a population mean (or average). (For example, estimating from a *sample* of items in inventory the average age of *all* the items in inventory, or estimating from a *sample* of components produced by a certain firm the average useful life of *all* the components produced by that firm.)

Chapter 6: Interval Estimation for Proportions, Mean Differences, and Proportion Differences

Here the methods of Chapter 5 are extended to address three additional estimation cases: building confidence interval estimates of a *population proportion* (e.g., the proportion of items in a population that are large, or over 30, or flawed, etc.); building confidence interval estimates of the *difference between the means* of two populations (e.g., the difference between the average recovery time for a test group of patients and the average recovery time for a control group of patients); and building confidence interval estimates of the *difference between two population proportions* (e.g., the difference between the proportion of Democrats who support a temporary budget freeze and the proportion of Republicans who favor such a freeze).

Chapter 7: Statistical Hypothesis Testing: Hypothesis Tests for a Population Mean

Chapter 7 introduces the second side of statistical inference, statistical **hypothesis testing**. In hypothesis testing, a statement is first made about a population of values. (For example, "The average life of all tires produced by company XYZ is 40,000 miles," or, "The average years of experience for all employees of the firm is 6.5 years.") A sample is then selected to help decide whether or not the statement is true (or "believable"). The trick, as we'll see, is to balance the chances of reaching wrong conclusions: the chance of believing that the statement is true when it's actually false versus the chance of believing the statement is false when it's actually true.

Chapter 8: Hypothesis Testing for Proportions, Mean Differences, and Proportion Differences

Here hypothesis testing principles are applied to the three cases listed in the chapter title.

Chapter 9: Regression Analysis

Chapter 9 introduces regression analysis, a statistical procedure commonly used to explore the relationship that may exist between certain key variables in a broad range of business and nonbusiness situations. For example, regression analysis might be used to investigate the possible relationship between the amount a firm spends on advertising and the firm's annual sales revenue; or the possible relationship between the number of inspections performed on units assembled by a company's production people and the number of units returned by customers as unsatisfactory; or the possible relationship between a baseball team's average player salary and its win-loss record. The application possibilities are almost unlimited.

1.6 THE NATURE OF DATA: LEVELS OF MEASUREMENT

One final note. Whenever data is collected, whether from historical records, casual observation, or controlled experimentation, the process of measurement—assigning proper values to observed phenomena—is involved. And whenever the process of measurement is involved, the issue of **levels of measurement** presents itself. Since most statistical techniques are suited to data measured only at certain levels, any user of statistical analysis should be aware of the level of measurement involved in his or her data. We can, in fact, describe four distinct measurement levels.

Nominal Data

Nominal data represents the lowest level of measurement. With nominal data, each value serves strictly as a label or a name. For example, a country of origin data set might include the possible "values" France (which might be designated country 1), the United States (country 2), Japan (country 3), and so on. With this sort of data, there's no natural ordering of values; the "value" of France (1) is no larger or smaller than the "value" of Japan (3). Each value is just an identifier.

Ordinal Data

Ordinal data represents a step up on the measurement scale. Here, in contrast to the nominal data case, values can be meaningfully rank-ordered. For example, we might create an ordinal data set by asking an audience to rank-order five previewed television shows. Or we might want to rank all-time great baseball players. In most such cases, a rank (or value) of 1 would be considered "higher" than (or superior to) a rank of 2, a rank of 2 would be considered higher than a rank of 3, and so on. It's important to note, however, that with ordinal data, even though rank-ordering is possible, measuring the precise distance between successive ranks is normally difficult or impossible. (A rank of 3 is higher than a rank of 4, but HOW MUCH higher?) Furthermore, there's no reason to believe that the distance between a number 1 ranking and a number 2 ranking is the same as the distance between a number 3 and a number 4 ranking. Nor is it clear that a rank of 4 is twice as "bad" as a rank of 2.

Interval Data

Interval data allows not only a rank-ordering of values, but also shows standardized, well-defined distances between successive values on the measurement scale. Temperatures measured on the Fahrenheit scale are frequently cited as an example of interval data. A 40° day is exactly 10° warmer than a 30° day, the same distance that separates a 70° day from a 60° day. (However, it would be incorrect to say that a 60° day is precisely twice as warm as a 30° day. This point brings us to the next category.)

Ratio Data

At the highest level of measurement, **ratio** data has all the properties of interval data, plus a so-called natural "zero" point, allowing for ratio comparisons. For example, a data set consisting of the heights of family members would show data measured on a ratio scale. We could say Bob (72 inches) is twice as tall as Jim (36 inches) but only two-thirds as tall as Peter (108 inches). Not only are there measurable and meaningful differences between the heights, but ratio comparisons also have meaning. (The natural "zero" point on the height scale indicates a complete absence of height. It's this fact that allows for effective ratio comparisons. Compare this to the case of Fahrenheit temperatures, where zero degrees clearly doesn't indicate a complete absence of heat.)

As a general rule, any statistical procedure that is appropriate to data at one level of measurement will be appropriate at any higher level of measurement, but not necessarily at lower levels. For example, computing the median value to describe the center of a data set—something we'll discuss in Chapter 2—makes sense for data at the ordinal level of measurement and above (i.e., for interval and ratio data), but not for nominal data. The inferential procedures described in Chapters 5 through 9 require at least interval data.

CHAPTER SUMMARY

Chapter 1 provided an introduction to the study of probability and statistics. Statistics was seen generally as the means to convert data into information. Descriptive statistics offers ways of summarizing, describing, and presenting data effectively. Inferential statistics (sampling theory) provides the means to draw conclusions (inferences) about large data sets (populations) based on the information acquired from smaller, representative subsets (samples). Probability theory, the study of risk or uncertainty, was seen as a link between descriptive and inferential statistics.

To put statistical methods in a broader perspective, we discussed two analytic views of the world and asserted that the tools of probability and statistics were suited to a probabilistic, rather than a deterministic, world view.

Finally, the levels of measurement involved in the collection and analysis of data were discussed, since various statistical procedures may require specific levels of measurement. Nominal (name), ordinal (rank), interval (showing standardized distances between values), and ratio (with a natural zero point) data were described and differentiated.

KEY TERMS

- Descriptive statistics
- Hypothesis testing
- Inferential (or inductive) statistics
- Interval estimation
- Levels of measurement
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Parameter
- Probability theory
- Statistic
- Statistics