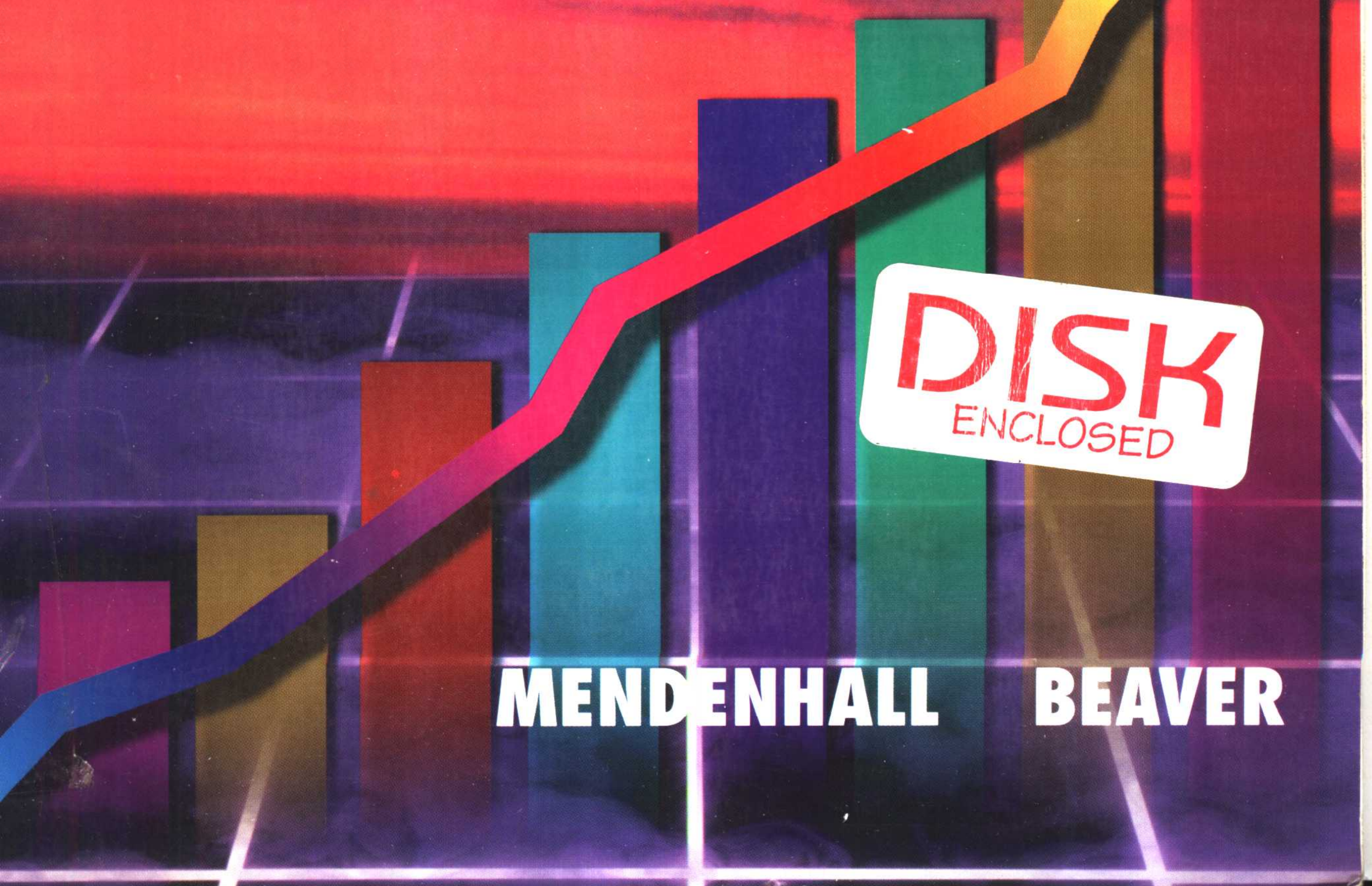
A *BRIEF* COURSE IN BUSINESS STATISTICS



MENDENHALL BEAVER

A Brief Course in BUSINESS STATISTICS

William Mendenhall
University of Florida, Emeritus

Robert J. Beaver
University of California, Riverside

Duxbury Press
An Imprint of Wadsworth Publishing Company

I(T)P™ *An International Thomsom Publishing Company*

Belmont • Albany • Bonn • Boston • Cincinnati • Detroit • London • Madrid • Melbourne
Mexico City • New York • Paris • San Francisco • Singapore • Tokyo • Toronto • Washington

Editor: *Curt Hinrichs*
Assistant Editor: *Jennifer Burger*
Editorial Assistant: *Michelle O'Donnell*
Production Editor: *Sandra Craig*
Managing Designer: *Cloyce Wall*
Print Buyer: *Randy Hurst*
Permissions Editor: *Peggy Meehan*
Cover Designer: *Juan Vargas, Vargas/Williams/Design*
Marketing Manager: *Joanne Terhaar*
Printer: *R. R. Donnelley & Sons/Crawfordsville*

I(T)P The ITP logo is a trademark under license

Printed in the United States of America
1 2 3 4 5 6 7 8 9 10—01 00 99 98 97 96 95

Printed on acid-free recycled paper.

For more information, contact Wadsworth Publishing Company.

Wadsworth Publishing Company
10 Davis Drive
Belmont, California 94002, U.S.A.

International Thomson Publishing Europe
Berkshire House 168-173
High Holborn
London, WC1V 7AA, England

Thomas Nelson Australia
102 Dodds Street
South Melbourne 3205, Victoria, Australia

Nelson Canada
1120 Birchmount Road
Scarborough, Ontario, Canada M1K 5G4

International Thomson Editores
Campos Eliseos 385, Piso 7
Col. Polanco
11560 México D.F. México

International Thomson Publishing GmbH
Königswinterer Strasse 418
53227 Bonn, Germany

International Thomson Publishing Asia
221 Henderson Road
#05-10 Henderson Building
Singapore 0315

International Thomson Publishing Japan
Hirakawacho Kyowa Building, 3F
2-2-1 Hirakawacho
Chiyoda-ku, Tokyo 102, Japan

Library of Congress Cataloging-in-Publication Data

Mendenhall, William.
A brief course in business statistics / William Mendenhall, Robert J. Beaver.
p. cm.
Includes bibliographical references and index.
ISBN 0-534-25290-7
1. Commercial statistics. I. Beaver, Robert J. II. Title.
HF1017.M45 1994
519.5—dc20 94-17706

Preface

In recent years business statistics textbooks have grown longer and longer with increased topic coverage, while at the same time many business programs have reduced the number of hours devoted to statistics. To solve this problem, we have created *A Brief Course in Business Statistics,* which is tailored to the needs of a one-quarter or one-semester course in introductory business statistics.

Although many useful and important topics could have been included in this text, we believe that the topics selected for inclusion are the most basic and essential for an understanding of the subject. It is our hope that students who complete a course using this text will want to delve into more advanced topics. In fact, for courses that require additional statistical methods, we recommend the book upon which this briefer text is based, *A Course in Business Statistics,* Third Edition. The mathematics requirement for both texts is high school algebra.

Our aim in this book is to provide students with the background to appreciate the statistics reported in the media that affect our daily lives and to understand the role of statistics in making informed business decisions.

Approach

Rather than proving theorems and manipulating equations, this book approaches statistics by showing students how to solve problems in the presence of uncertainty—the situation confronting managers in the business environment every day. Throughout the text students are reminded of the objective of statistics: "to make inferences about a population from information contained in a sample drawn from that population, and to assess the reliability of the inference." Each chapter plays its role in helping students understand this objective: first by showing how populations and samples are described, then by laying the probabilistic foundation by which the reliability of an inference is measured, and finally by demonstrating a number of statistical procedures used in analyzing business data and the kinds of inferences that are possible using these procedures. As each chapter builds on the material in preceding chapters, learning is encouraged through repetition and development of concepts. In addition, typical examples from current newspaper and magazine articles demonstrate how the statistical techniques and procedures are applied. Throughout the text the vocabulary of statistics is carefully defined in context, and new terms are used repeatedly to solidify comprehension.

Organization and Coverage

The ordering of chapters and the material within chapters reflects the pedagogy we use in our own courses. We believe that Chapters 1 through 8 should be covered in the order presented. However, Chapter 10, concerning chi-square goodness-of-fit tests, can be covered in its present position, before Chapter 9, or after Chapter 12. Although some texts cover linear regression before the analysis of variance, we introduce the analysis of variance in Chapter 8 as a generalization of the two-sample procedures. This order helps students understand the computer printouts for regression in Chapter 12, which use the analysis of variance in testing for significant regression. This order also complements the use of Student's t-test in testing for significance of total or partial regression coefficients. Chapter 13, on quality control, can be covered any time after Chapter 8.

The Role of the Computer

The Minitab computer software package is integrated throughout the text as the primary statistical package for data analysis. The Statistical Analysis Systems (SAS) package is also used in those chapters that deal with regression. We chose Minitab because its simplicity makes it excellent for beginning statistics students. Students using other statistical software should have no difficulty with this if they focus on the important items on a printout since outputs from the various statistical packages generally include the same information.

Ancillaries and the Data Disk

An Instructor's Manual containing solutions to all problems in the text is available to instructors who adopt the book. A Partial Solutions Manual, which contains complete, worked-out solutions to approximately one-fourth of the exercises in the text, is for sale to students. Each new copy of the text includes a data disk that contains data of most problem sets in the text and two large data sets. These large data sets consist of the average salaries of instructional staff by rank and gender for 185 colleges, and measurements on the yield characteristics of broccoli as a function of ozone levels. The data on the $3\frac{1}{2}''$ MS-DOS disk is easily readable by most popular statistical software packages and spreadsheets; a Macintosh disk is also available.

Acknowledgments

We would like to thank the staff of Duxbury Press, especially our editor, Curt Hinrichs; Jennifer Burger, assistant editor; and Michelle O'Donnell, editorial assistant. We also thank the many reviewers and users of our texts and their students who have studied from them for their thoughtful and candid feedback. We extend a special note of gratitude to Barbara Beaver for her help in manuscript preparation and her diligence in preparing the solutions manuals that accompany this text.

William Mendenhall
Robert J. Beaver

Contents

6 *Sampling Distributions* *176*

7 *Estimation of Means and Proportions* *212*

8 *Tests of Hypotheses for Means and Proportions* 266

9 *The Analysis of Variance* 347

CHAPTER 1

What Is Statistics?

The purpose of this chapter is to identify the nature of statistics, its objective, and how it plays an important role in the sciences, in industry, and, ultimately, in our daily lives.

CASE STUDY

Reaping the Rewards of Refunds

Will court-enforced retail refunds make us rich? Not likely, according to the *Wall Street Journal* article (June 9, 1981), which reported on a court case stemming from a price-fixing indictment. According to the article, approximately 54,000 persons — customers of Saks Fifth Avenue, Bonwit Teller, and Bergdorf Goodman — during the period of 1968 to 1974, may be due refunds if they can resurrect the ancient records of their purchases and file the necessary applications. This case raises a serious question about this and other court-directed refund programs that are often the result of price-fixing indictments. Who profits, other than the attorneys involved in the lawsuit? Specifically, how many of the 54,000 customers eligible for refunds will actually find their receipts and be willing to submit the necessary applications? And how much of the settlement money will actually be distributed to those customers? To answer these questions, we would need to either interview all 54,000 customers who are potentially eligible — an impossible task — or select a sample from among the 54,000 and use statistical methods to estimate the number eligible for refunds and the amount of money that will eventually be refunded.

Our intention is not to answer the question "Who profits from refunding?" but to present a difficult problem that can be solved by the use of a very powerful tool, statistics. In this chapter we will describe the objective of statistics. Specifically, we will identify the types of problems that statistical methodology can solve and then explain how this valuable tool can be used to answer some practical questions. We will revisit the refunding case study in Section 1.5.

1.1 Illustrative Statistical Problems

What is statistics? How does it function? How does it help to solve certain practical problems? Rather than attempt a definition at this point, let us examine several problems that might come to the attention of the statistician. From these examples we can then select the essential elements of a statistical problem.

To predict the outcome of a national election, pollsters interview a predetermined number of people throughout the country and record their preferences. On the basis of this information a prediction is made. Similar problems are encountered in market research (What fraction of potential buyers prefer automobile brand A?); in sociology (What fraction of rural homes have electricity?); in industry (What fraction of items purchased, or produced, are defective?).

An auditor wants to determine the inventory of a large hospital. To count the number and value of each expendable and nonexpendable item in stock would not only be very costly but would also, because of the size of the task, be subject to error. To reduce the cost and obtain a reliable estimate of value, the auditor selects a sample of items from the list of the hospital's supplies and equipment, carefully counts the number of each item on hand, and records its value. The ratio of the total value of this sample of items to the total value shown in the hospital's records provides an estimate of the shrinkage due to theft, failure to record use of items, and so on. This shrinkage rate can then be applied to the total value of inventory shown in the hospital's records, thereby obtaining an estimate of the actual value of current inventory. How accurate is this estimate? How far might we expect the estimate to deviate from the actual value of the hospital's inventory?

The yield (production) of a chemical plant is dependent upon many factors. By observing these factors and the yield over a period of time, we can construct a prediction equation relating yield to the observed factors. As another example, the economist wants to develop prediction equations that will be useful in forecasting growth or some other measure of economic health, as a function of other variables. Similarly, a manager may want to predict the sales of a product as a function of advertising expenditure, number of salespersons employed, or various other variables that may be related to the company's sales.

How do we find a good prediction equation? If the equation is used to predict yield, the prediction will rarely equal the true yield; that is, the prediction will almost always be in error. Can we place a limit on the prediction error? Which factors are the most important in predicting yield?

In addition to being involved in prediction, statistics is concerned with decision making based on observed data. Consider the problem of determining the effectiveness of a new flu vaccine. For simplification, let us assume that ten people have received the new flu vaccine and are observed over a winter season. Of these ten, eight survive the winter without acquiring the flu. Is the vaccine effective?

Two different teaching techniques are used to present a subject to two groups of students of comparable ability. At the end of the instructional period a measure of achievement is obtained for each group. On the basis of this information we ask: Do the data present sufficient evidence to indicate that one method produces, on the average, higher student achievement?

Consider the inspection of items purchased for a manufacturing plant. On the basis of such an inspection each lot of incoming goods must be either accepted or rejected and returned to the supplier. The inspection might involve drawing a sample of ten items from each lot and recording the number of defectives. The decision to accept or reject the lot could then be based on the number of defective items observed.

A company manufacturing complex electronic equipment produces some systems that function properly but also some that, for unknown reasons, do not. What makes good systems good and bad systems bad? To answer this question, we might make certain internal measurements on a system to find important factors that differentiate between an acceptable and an unacceptable product. From a sample of good and bad systems, data could then be collected that might shed light on the fundamental design or on production variables that affect system quality.

1.2 The Population and the Sample

The examples we have cited vary in nature and complexity, but each involves prediction or decision making. In addition, each of these examples involves sampling. A specified number of items (objects or bits of information)—a **sample**—is drawn from a much larger body of data, which we call the **population**. The pollster draws a sample of opinions (those interviewed) from the statistical population, which is the set of opinions corresponding to all the eligible voters in the country. In predicting the fraction of potential buyers who prefer automobile brand A, we assume that those interviewed yield a representative sample of the population of all potential automobile buyers. The sample for the flu vaccine experiment consists of observations made on the ten individuals receiving the vaccine. The sample is presumably representative of data pertinent to a much larger body of people—the population—who could have received the vaccine.

Which is of primary interest, the sample or the population? In all of the examples given above, we are primarily interested in the population. We cannot interview all the people in the United States; therefore, we must predict their behavior on the basis of information we obtain from a representative sample. Similarly, it is practically impossible to give all possible users a flu vaccine. The manufacturer of the drug is interested in its effectiveness to prevent the flu in the purchasing public (the population). They can predict this effectiveness from information extracted from the sample. Therefore, the sample may be of immediate interest but we are primarily interested in describing the population from which the sample is drawn.

DEFINITION

The *population* is the set representing all measurements of interest to the investigator.

DEFINITION

A *sample* is a subset of measurements selected from the population of interest.

Most people give the word *sample* two meanings. They refer to it as the set of objects on which measurements are to be taken, or they use it to refer to the objects themselves. A similar double use could be made of the word *population*.

For example, we read in the newspapers that a Gallup Poll was based on a sample of 1823 people. In this use of the word *sample*, the objects selected are obviously people. Presumably, each person is interviewed on a particular question and that person's response represents a single item of data. The collection of data corresponding to the people represents a sample of data.

In a study of sample survey methods we must distinguish between the objects measured and the measurements themselves. To experimenters, the objects measured are called **experimental units**. The sample survey statistician calls them **elements of the sample**.

To avoid a proliferation of terms, we will use the word *sample* in its everyday meaning. Most of the time, we will be referring to the set of measurements made on the experimental units (elements of the sample). If occasionally we use the term to refer to a collection of experimental units, the context of the discussion will clarify the meaning.

EXAMPLE 1.1

A fast-food company wishes to know how much money Americans, say age 16 and over, will spend on fast foods during the first week of June. Describe the population of data of interest to the company. Explain how the company might acquire the information it desires.

SOLUTION The information that the fast-food company wishes to acquire is associated with a population of measurements—one measurement for a week's expenditure on fast foods for each of the many millions of Americans in the age group 16 and over. These measurements will vary.

Some persons will purchase no fast foods during the week; others will purchase varying amounts. To acquire information on this vast population of measurements, the fast-food company likely will hire a marketing research organization to sample the population. From the sample measurements the marketing organization will estimate the average amount spent by all Americans age 16 and over, will estimate the proportion of all Americans who buy fast foods, and will answer other similar questions posed by the fast-food company. Thus the marketing organization will use information contained in the sample to infer the nature of the population of fast-food expenditures during the first week of June.

1.3 The Essential Elements of a Statistical Problem

You can see from the preceding discussion that statistics is concerned with describing a data set called a population. In some rare instances, such as the United States census, the population will be stored in a computer, and the statistical problem is to describe and extract information from a large mass of data. The branch of statistics concerned with this type of problem is called **descriptive statistics**. Usually, however, the population is unavailable. Either observing and recording every single member of the population is too costly (as in the case of the hospital audit), or the population is conceptual—that is, the set of measurements (such as the daily yields of a chemical plant over the next two years) exists in our minds but is not actually available. Using a sample from a population, we attempt to deduce the population's nature by using the branch of statistics known as **inferential statistics**. Each of the examples described in Section 1.1 represents a problem in inferential statistics. All have the same objective.

The objective of inferential statistics

> The objective of inferential statistics is to make inferences (predictions, decisions) about a population based on information contained in a sample.

How will we achieve this objective? We will find that every statistical problem contains five elements. The first and foremost of these is a clear **specification of the question** to be answered and **identification of the population** of data related to it.

The second element of a statistical problem is the decision about how the sample will be selected. This element, called the **design of the experiment** or the **sampling procedure**, is important because data cost money and time. In fact, it is not unusual for an experiment or a statistical survey to cost \$50,000 to \$500,000, and the costs of many biological or technological experiments can run into the millions. What do these experiments and surveys produce? Numbers on a sheet of paper or, in brief, information. Therefore, planning the experiment is important. Including too many observations in the sample is often costly and wasteful; including too few is also unsatisfactory. Most importantly, you will learn that the method used to collect the sample will often affect the amount of information per observation. A good sampling design can sometimes reduce the costs of data collection to one-tenth or as little as one-hundredth of the cost of another sampling design.

The third element of a statistical problem involves the **analysis of the sample data**. No matter how much information the data contain about the practical question, you must use an appropriate method of data analysis to extract the desired information from the data.

The fourth element of a statistical problem is the use of the sample data to make an **inference** about the population. As you will learn, many different procedures can be used to make an estimate or decision about some characteristic of a population or to predict the value of some member of the population. For example, two different methods may be available to estimate consumer response to an advertising campaign, but one procedure may be much more accurate than another. Therefore, you will want to use the best inference-making procedure when you use sample data to make an estimate or decision about a population, or a prediction about some member of a population.

The final element of a statistical problem identifies what is perhaps the most important contribution of statistics to inference making. It answers the question, how good is the inference? To illustrate, suppose you manage a small manufacturing concern. You arrange for an agency to conduct a statistical survey for you, and it estimates that your company's product will gain 34% of the market this year. How much faith can you place in this estimate? You will quickly realize that you are lacking some important information. Of what value is the estimate without a measure of its reliability? Is the estimate accurate to within 1%, 5%, or 20%? Is it reliable enough to be used in setting production goals? Statistical estimation, decision making, and prediction procedures enable you to calculate a **measure of goodness** for every inference. Consequently, in a practical inference-making situation, every inference should be accompanied by a measure that tells you how much faith you can place in the inference.

To summarize, a statistical problem involves the following.

1. A clear definition of the objective of the experiment and the pertinent population.
2. The design of the experiment or sampling procedure.