INTRODUCTORY STATISTICS FOR BUSINESS AND ECONOMICS

Richard K. Miller

Introductory Statistics for Business and Economics

RICHARD K. MILLER
Washington and Jefferson College

ST. MARTIN'S PRESS
New York

Library of Congress Catalog Card Number: 80-51065

Manufactured in the United States of America.
43210
fedcba
For information, write St. Martin's Press, Inc.,
175 Fifth Avenue, New York, N.Y. 10010

cover design: Notovitz & Perrault

typography: Leon Bolognese

ISBN: 0-312-43451-0

ACKNOWLEDGMENTS

"The Standard Normal Distribution" reprinted with permission of Macmillan Publishing Co., Inc. from "Areas under the Normal Distribution" in *Elements of Econometrics* by Jan Kmenta. Copyright © 1971 by Macmillan Publishing Co., Inc.

"The t Distribution" taken with permission from Table III of Ronald A. Fisher and Frank Yates: *Statistical Tables for Biological, Agricultural and Medical Research,* 6th edition, 1974, published by Longman Group Ltd., London (previously published by Oliver and Boyd, Edinburgh), and reprinted by permission of the authors and publishers.

"The χ^2 Distribution" from Table 8, "Percentage Points of the χ^2 Distribution" in E. S. Pearson and H. O. Hartley, editors, *Biometrika Tables for Statisticians,* volume I, 3rd edition (London: Cambridge University Press for the Biometrika Trustees, 1966). Reprinted with permission.

"The F Distribution" ("Critical Values for the Upper .05 Area" and "Critical Values for the Upper .01 Area") from Table 18, "Upper 5% Points and Upper 1% Points of the F Distribution" in E. S. Pearson and H. O. Hartley, editors, *Biometrika Tables for Statisticians,* volume I, 3rd edition (London: Cambridge University Press for the Biometrika Trustees, 1966). Reprinted with permission.

"Table of Random Numbers" from The RAND Corporation, *A Million Random Digits with 100,000 Normal Deviates* (Glencoe, Ill.: The Free Press, 1955), p. 337. Reprinted with permission.

Preface

Introductory Statistics for Business and Economics is the by-product of a Business Statistics course that I have taught at Washington and Jefferson College for the past nine years. During that time, I have become aware of the large number of economics and business administration majors who view exposure to mathematics of *any* kind with a pronounced degree of trepidation and defeatism. In spite of a proliferation of statistics texts, such fears persist.

The Mathematical Association of America recently reported that introductory statistics classes often "are too technique-oriented, overemphasizing computation and underemphasizing the fundamental ideas underlying statistical reasoning."* This indictment can also be extended to many of the books associated with these courses. Although a host of texts carry the claim that "only high school algebra is required," some students are still overwhelmed by a frustrating array of symbols and equations. Other students are doubtful whether statistical procedures have a practical use.

I have written this text with two objectives in mind: (1) to illustrate how statistical analysis can be applied to "real-world" situations, and (2) to discuss the basic theory of statistics without relying on complex symbols and advanced mathematics. To accomplish the first objective, I have included an extensive collection of worked-out business and economics examples in the body of the text, as well as a variety of end-of-chapter exercises. The second objective is attained through the presentation of numerous tables and diagrams and by a clear step-by-step approach to proving key principles.

I feel that most introductory statistics courses, regardless of the type of enrollment, share a common core of subject matter. This basic material is presented in Chapters One through Eleven. Following this core, the direction of individual courses is determined by both the preferences of the instructor and the characteristics of the students. Chapters Twelve through Twenty are especially relevant for majors in business and economics. I believe that instructors of such students will find that this text offers enough "additional" topics to satisfy divergent concepts of what the introductory business statistics course should entail.

* *A Compendium of CUPM Recommendations,* vol. II, 1975, p. 475.

The text can be used in either a one- or two-semester program. The accompanying diagram indicates the organization.

	Descriptive Statistics	*Probability*	*Statistical Inference*
Core Chapters	2, 3	4, 5, 6	7, 8, 9, 10, 11
Additional Chapters	16, 19		12, 13, 14, 15, 17, 18, 20

I have been fortunate in being able to develop *Introductory Statistics for Business and Economics* with the help of St. Martin's Press. Not only have I enjoyed my friendships with the staff, but I have been impressed with their dedication and professionalism. I particularly want to express my appreciation to Bertrand W. Lummus, senior editor; Carolyn Eggleston, manuscript editor; Edward B. Cone, managing editor; and Thomas V. Broadbent, director of the college department.

The text was also enhanced by the comments of many reviewers. Dr. Ronald S. Koot of Pennsylvania State University was exceptionally helpful and articulate.

Finally, 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 Table III from their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974).

RICHARD K. MILLER

Contents

1 Introduction

RELEVANCE OF STATISTICS

Because we are so persistently exposed to numerical information in our daily lives, almost everybody is a player in the "numbers game" called *statistics*. To the man on the street, the words "statistics" and "numbers" imply the same thing. There are, however, distinctions, as we shall see. Techniques for obtaining, classifying, and evaluating quantitative information fall under the heading of "statistics." Thus statistical methods are used in the construction of Census Bureau questionnaires, the selection of recipients of these questionnaires, and the tabular display of responses. We will shortly see, moreover, that "averages" are called "statistics" under certain circumstances. Throughout this text, we will illustrate the many uses of statistics in business and economics. Additionally, we will apply statistical methods to problems from many other fields, for example, politics.

Preenrollment anxiety over a class in introductory statistics may be reinforced early in the semester, perhaps when

$$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$$

is first written on the blackboard. There is no escaping the fact that formulas play a large role in statistics. We will try, though, to develop the rationale or logic behind such relationships, which hopefully will make the formulas appear less forbidding. In addition, the mathematical equipment that is needed is intentionally restricted to algebra and/or arithmetic.

Newspapers and weekly periodicals have been giving increasing attention to statistical studies. The Gallup, Harris, and Roper polls of American opinion are printed regularly. Debates in the news media as to whether cigarette smokers are more likely than nonsmokers to get particular diseases or whether various working conditions adversely affect health rely heavily on statistical evidence.

Students who complete an introductory statistics course should have acquired the ability to communicate quantitative information effectively and efficiently. They should be able to identify types of real-life situations in which statistics and statistical techniques are useful. Furthermore, they ought to understand that statistical analysis, in spite of its theoretical foundation and sophisticated form, is *not without its limitations.*

Some students complain, "Why must I take Statistics?" We hope to answer this by integrating the methods of statistics with interests common to business people and economists. To emphasize the text's concern with "applications," Chapter One will outline a number of problems that can be dealt with statistically. When you look over these examples, imagine how you might approach them now. Quite possibly, you would rely on intuition. Before too long, you'll begin to realize that the procedures explained in the following chapters have advantages over "hit-and-miss," "seat-of-the-pants," or "rule-of-thumb" attempts at problem solving.

DIMENSIONS OF STATISTICS

The major topics discussed in this text are descriptive statistics (Chapters Two and Three), probability (Chapters Four through Six), statistical inference (Chapters Seven through Fifteen, and Twenty), linear regression (Chapters Sixteen through Eighteen), and time series (Chapter Nineteen). Before we illustrate their principal uses, it is necessary to introduce some important terminology.

The information collected for any statistical study is associated with a *variable of interest.* Such a variable could be the salaries of corporation presidents, the advertising expenses of tobacco companies, or the debt/equity ratios of firms in the steel industry. A single piece of information (e.g., the debt/equity ratio of U.S. Steel Corp.) is called either an *item* or an *observation.* These items are the researcher's *data.* Table 1–1 illustrates these concepts.

If the six courses listed in Table 1–1 are the only business administration courses scheduled for the fall semester, then every possible item relevant to the "enrollments" variable is given in the table. Under these circumstances, we say that the column headed "Students Enrolled" (40, 20, 30, 35, 15, 40) constitutes a *statistical population.*

TABLE 1–1
Enrollments in business administration courses taught at College T during the Fall 1980 semester — VARIABLE OF INTEREST

Course	Students Enrolled
Item #1: Business Statistics	40
Item #2: Cost Accounting	20
Item #3: Management	30
Item #4: Marketing	35
Item #5: Labor Relations	15
Item #6: Corporate Finance	40

(Students Enrolled values — DATA)

Researchers often find that they can't obtain observations for all of a particular population. Among other reasons, this task can be (1) quite expensive, (2) very time consuming, or (3) physically impossible. Practical considerations, therefore, may force researchers to examine only part of a population. We refer to parts or subsets of populations as *samples*. Table 1–2 breaks the enrollment data in Table 1–1 into three samples.

TABLE 1–2

Sample One (items 1, 2, and 4)	*Sample Two (items 2, 4, 5, and 6)*	*Sample Three (items 4 and 6)*
40	20	35
20	35	40
35	15	
	40	

How we define the variable of interest determines whether the data represent a population or a sample. Even though enrollments in the six courses in Table 1–1 form the population of "enrollments in business administration," they are just a sample of "enrollments in business administration, economics, and mathematics."

Numbers computed to *describe* a population or a sample are called *summary measures*. When summary measures derive from population data, they are said to be *parameters*. A *statistic* is a summary measure obtained from sample observations. The average enrollment per course in Table 1–1 is 30 students. In a sense, 30 "summarizes" the information supplied by the six items. Since the items form a population, 30 is a parameter. If the variable of interest is redefined as "enrollments in business administration, economics, and mathematics," the six items then constitute a sample and 30 becomes a statistic.

Parameters offer insights into populations by showing what is "average" or "typical." However, it is easier to collect data on samples than on populations. This raises a critical question: "How can researchers use a statistic to estimate an unknown parameter?" We might ask, for example, "If 100 of Company J's 6,000 employees are asked to state their feelings about working 10 hours a day for four days versus 8 hours a day for five days, will the responses be a good indication of the entire plant's preferences?"

We will first direct our attention to calculating summary measures. At the same time, we will suggest ways of displaying data. Determining parameters and statistics, as well as presenting data in an informative fashion, are objectives of *descriptive statistics.* About 40 or 50 years ago, introductory statistics courses were dominated by this topic. A descriptive technique, putting data in a table, is applied in the following example.

In the early 1970s, Senator Philip Hart of Michigan introduced an industrial reorganization bill that would have given Congress the power to break up industries composed of a few giant firms. Imagine that as part of its debate over this bill, a congressional subcommittee asked an economist to investigate whether there was a predictable relationship between corporate size and profitability. The economist believed that members of the subcommittee would be curious about the size of the firms used in her study. Consequently, she passed out Table 1–3 before testifying.

TABLE 1–3

Company Size (assets in $ billions)	*Number of Companies*
0–0.9	12
1.0–1.9	15
2.0–2.9	6
3.0–3.9	4
4.0 and over	3
	40

The central focus of statistics today is analysis of sample data for the purpose of gaining some understanding of the population from which the sample was drawn. This involves using methods of *statistical inference.* These methods enable researchers to *estimate* parameter values from sample statistics. The widely reported national unemployment rate is an estimate of a parameter.

The Bureau of Labor Statistics contacts roughly 47,000 households (approximately 100,000 people) every month. Any individual 16 years of age or older is said to be "employed" if he or she is currently working part-time or full-time. People who are "unemployed" are looking for a job, whereas someone "not in the labor force" is neither working nor seeking work. The labor force (i.e., the employed plus the unemployed) numbers over 100 million people. The statistic "unemployment rate for the sample" becomes an estimate of the parameter "unemployment rate for the country."

During presidential campaigns, the prevailing unemployment rate is frequently an issue. If the rate is high, the incumbent will argue that his administration was plagued by a series of unfortunate economic conditions and that steps are being taken to bring the rate down. The challenger will say that the high rate is just another result of the incumbent's ineffectiveness.

Although the unemployment rate is based on a sample of less than 1 percent of the labor force, presidential candidates (and others) generally will not attack the procedures used to derive it. There is a reason for this reluctance. Before a sample is selected, researchers can objectively compute the probability that their estimate will be near the "true" parameter value. When the Secretary of Labor maintains, "I am quite confident that the June unemployment rate was close to 5.8 percent," his "confidence" arises from the estimation technique itself, not from a boastful personality.

Besides estimation, statistical inference also includes *hypothesis testing.* In testing hypotheses, researchers tentatively assign a value to a parameter and then select a sample. By virtue of the characteristics of the sample, the hypothesized parameter value will either seem (1) reasonable or (2) doubtful. The researchers establish the criterion or criteria that will allow them to choose between 1 and 2. In the following example, the hypothesis "95 percent of all cars have been inspected" is subjected to a test.

Drivers in a particular state must have their cars inspected by a certified mechanic twice a year. The deadline for the current inspection period is March 31. Because the weather has been severe for many weeks, some people could not get to an inspection station. Should the deadline be extended? The Secretary of Transportation feels that if 95 percent of all cars have been inspected, there is no need to grant an extension. To determine whether this is the case, he has 1,000 of the state's registered car owners contacted. Only 90 percent of the cars in this sample were inspected. The sample "evidence" makes the secretary skeptical of the 95 percent parameter value. As a consequence, he declares an extension.

Notice that sample summary measures are involved in the inference process. *Probability theory,* or the "nature of chance," is the bridge which connects descriptive statistics and statistical inference. How we evaluate "chance" often influences our behavior, as the next example illustrates.

A handful of executives at Corporation Q play poker every Thursday night. Ted Burns, a new employee, is invited to join the poker group. Burns wins his first hand by drawing four aces. After a few executives joke about "beginner's luck," the second hand is dealt. When Burns again draws four aces, the laughing stops and one player calls him a cheater.

The group changed its attitude because the "chance" of getting two consecutive hands of four aces in a fair game is very slight; that is, two consecutive hands of four aces is *possible but improbable.* The players concluded that this outcome was "more probable" when a game was rigged.

The decision to extend the inspection deadline in the preceding example followed the same kind of thinking. Even if the parameter were 95 percent, the secretary knew that a sample having a 90 percent inspection rate could be selected. He felt, though, that a 90 percent sample outcome was "more probable" when the parameter was less than 95 percent.

Knowledge of probability theory is also essential for solving a problem such as the following.

Engineers at M. W. Electronics have put all the equipment in perfect order. Experience, nonetheless, suggests that 50 of the next 1,000 transistor components that are produced will be defective. Ms. Smith is going to receive a shipment of five transistor components. Management wants to compute the probability of this shipment's containing three defective parts. The company calculates that only 1 out of every 1,000 shipments of 5 transistor components will have 3 defectives. Therefore, management doesn't expect Smith to get three defectives.

Table 1–4 clarifies the distinction between estimation and hypothesis testing. It points out, moreover, the role that probability plays in making inferences.

In the next example, statistical inference is explicitly linked to managerial decision making. The scenario reveals that sampling can have disadvantages as well as advantages and shows the importance of probability theory.

A manufacturer produces 1,000 jar lids an hour. At the end of each hour, he must determine whether the machinery is cutting the lids to specifications. If specifications are not met, further production is curtailed until the equipment is fixed.

The manufacturer needs a strategy for detecting a malfunction. The equipment can be checked directly through inspection or indirectly by looking at all or some of the lids. Inspecting the equipment requires a substantial amount of time, during which the machines are shut down and no jar lids produced—a costly procedure. On the other hand, if the lids are defective, resources and manpower will have been wasted.

Rather than inspecting the equipment every hour, the manu-