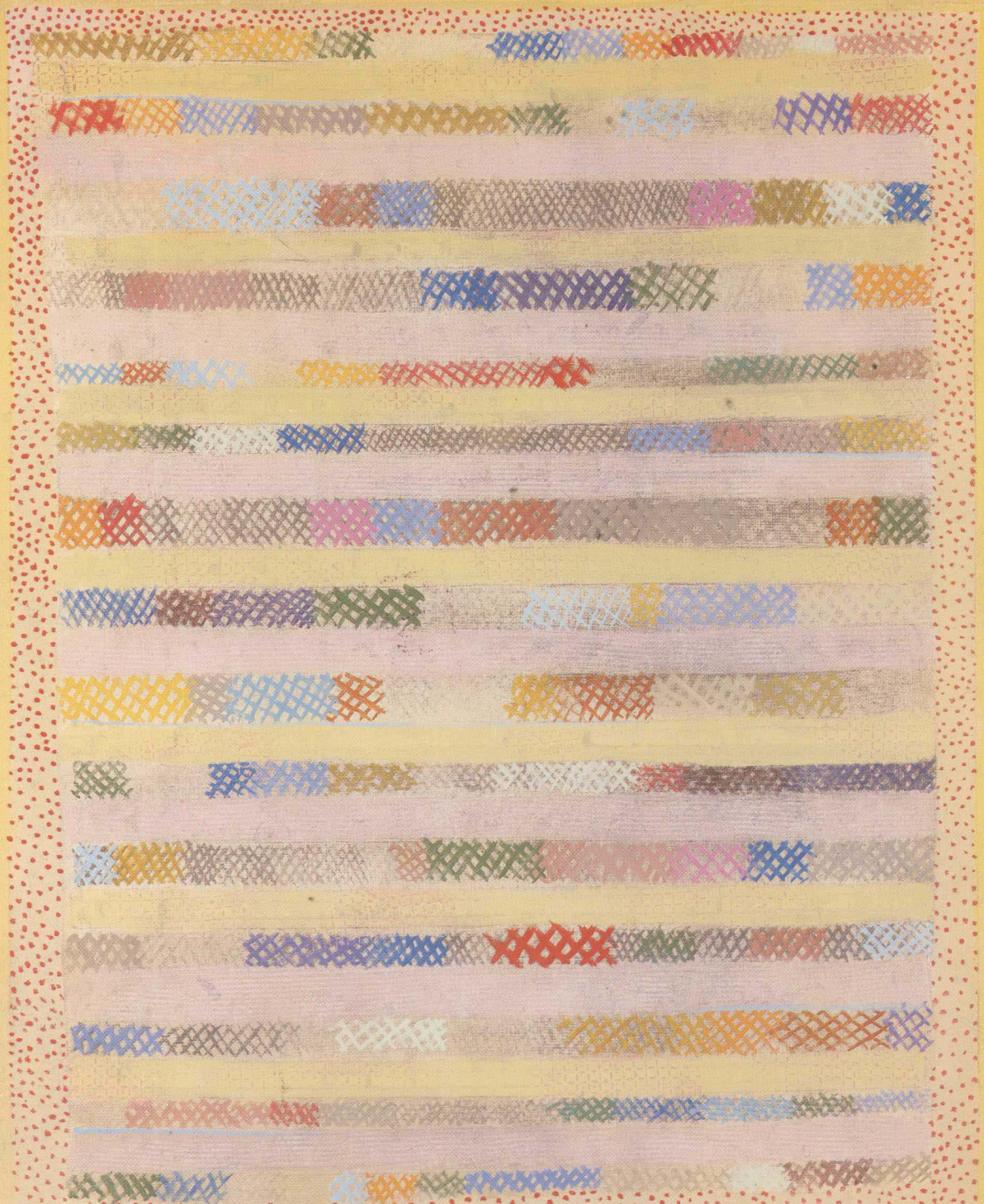
STATISTICS

THIRD EDITION

JAMES T. McCLAVE AND FRANK H. DIETRICH II



Third Edition

Statistics

James T. McClave
University of Florida

Frank H. Dietrich, II
Northern Kentucky University

Dellen Publishing Company
San Francisco, California

Collier Macmillan Publishers
London

divisions of Macmillan, Inc.

On the cover: The work on the cover is an acrylic painting titled "Homage to Louis Sullivan and Frank Lloyd Wright," 1983, by Robert Natkin. His work is in many private and public collections, among them the Museum of Modern Art, the Guggenheim Museum, the Metropolitan Museum of Art, and the Los Angeles County Museum. Natkin's work may be seen at the Tortue Gallery in Santa Monica, California, and at the Gimpel and Weitzenhoffer Gallery in New York City.

Permissions: Dellen Publishing Company
400 Pacific Avenue
San Francisco, California 94133

Orders: Dellen Publishing Company
c/o Macmillan Publishing Company
Front and Brown Streets
Riverside, New Jersey 08075

Collier Macmillan Canada, Inc.

Library of Congress Cataloging in Publication Data

McClave, James T.
Statistics.

Includes index.
1. Statistics. I. Dietrich, Frank H. II. Title.
QA276.12.M4 1985 519.5 84-28645
ISBN 0-02-378760-0 (Macmillan)

Printing: 3 4 5 6 7 8 Year: 5 6 7 8 9 0

ISBN 0-02-378760-0

Preface

The third edition of *Statistics* maintains the same objectives as the earlier editions, namely to introduce students to the basic concepts of statistics and to show them how these concepts can be used in making inferences from experimental data and from sample surveys. The text is designed so that the early chapters can be used for a one-quarter (or a one-semester) introductory course for all undergraduates. The remainder of the text can be used as a second-quarter (semester) follow-up course, with emphasis on special applications such as analysis of variance, regression analysis, or one of the other methodologies included in the later chapters. As in the second edition, we have maintained a unified approach to the subject, attempting both to provide an overall picture of statistics and its role in business and the sciences, and to provide the student with some methodology that will be relevant and useful in other college courses and in subsequent fields of employment. We have also maintained the same level of presentation in this edition.

In addition to making important changes in wording that improve the readability of the text, the following major changes have been made:

Chapter 2: Methods for Describing Sets of Data Two new sections have been added to this chapter. Section 2.2 explains how to construct and interpret stem and leaf displays. The data sets used to illustrate the methods of constructing stem and leaf displays and histograms are based on real data selected from newspapers and scientific journals. Section 2.8 explains how to use z-scores and box plots to identify outliers. Then this concept is used to introduce the notion of a rare event and a brief introduction to statistical inference in Section 2.9.

Chapter 5: Continuous Random Variables In the second edition, all the examples in Section 5.2 were concerned with finding areas under the normal curve between two values of the standard normal random variable z or two values of a normal random variable x. We have added an example that deals with the reverse problem, finding a value z_0 corresponding to an area under the normal curve.

Chapter 6: Sampling Distributions In Section 6.2, we have added a definition for a point estimator.

Chapter 7: Estimation and Tests of Hypotheses: Single Sample This edition attempts to provide a better explanation of the role of the Type I and Type II errors in testing hypotheses. The new Case Study 7.2 explains the role of hypothesis testing in the formulation of computer security systems. The relevance of Type I and Type II errors is explained in terms of the probabilities of rejecting authorized users and accepting unauthorized users. In addition, new Example 7.3 provides a second example of the computation of the probability β of making a Type II error.

Chapter 9: Analysis of Variance: Comparing More Than Two Means This chapter has been greatly expanded by adding two new sections, many new examples, and corresponding exercises. New Section 9.3 explains how to perform an analysis of variance for a two-factor factorial experiment. Particular stress is placed on the importance of detecting factor interactions. New Section 9.4 presents Tukey's method for making multiple comparisons about a set of population means.

Applications Particular effort was made to add exercises that are based on real-life situations, data sets, and/or research results reported by the news media or in research journals. A total of 184 new conceptual exercises of this type have been added in this edition.

Case Studies Eight new case studies have been added in this edition to place greater emphasis on the relevance of statistics to problem solving in the real world.

The third edition of *Statistics* still retains the most important features of the second edition. The material is presented in a manner that permits flexibility in the amount of time devoted to particular topics. Sections that are not prerequisite to succeeding sections and chapters are marked "(Optional)." We have included several features in this book that will make it different from most introductory statistics texts currently available. These features, which assist the student in achieving an overview of statistics and an understanding of its relevance in the solution of real problems, are as follows:

Case Studies (See the list of 30 case studies on page xvii.) Many important concepts are emphasized by the inclusion of case studies, which consist of brief summaries of actual applications of statistical concepts and are often drawn directly from the research literature. These case studies allow the student to see applications of important statistical concepts immediately after their introduction. The case studies also help to answer by example the often-asked questions, "Why should I study statistics? Of what relevance is statistics to my program?" Finally, the case studies constantly remind the student that each concept is related to the dominant theme—statistical inference.

The Use of Examples as a Teaching Device We have introduced and illustrated almost all new ideas by examples. Our belief is that most students will better understand definitions, generalizations, and abstractions *after* seeing an application. In most sections, an introductory example is followed by a general discussion of the procedures and techniques, and then a second example is presented to solidify the understanding of the concepts.

A Simple, Clear Style We have tried to achieve a simple and clear writing style. Subjects that are tangential to our objective have been avoided, even though

some may be of academic interest to those well-versed in statistics. We have not taken an encyclopedic approach in the presentation of material.

Many Exercises—Labeled by Type The text has a large number (almost 1,200) of exercises illustrating applications in almost all areas of research. However, we believe that many students have trouble learning the mechanics of statistical techniques when problems are all couched in terms of realistic applications—the concept becomes lost in the words. Thus, the exercises at the ends of all sections are divided into two parts:

1. **Learning the mechanics** These exercises are intended to be straightforward applications of the new concepts. They are introduced in a few words and are unhampered by a barrage of background information designed to make them "practical," but which often detracts from instructional objectives. Thus, with a minimum of labor, the student can recheck his or her ability to comprehend a concept or a definition.
2. **Applying the concepts** The mechanical exercises described above are followed by realistic exercises that allow the student to see applications of statistics across a broad spectrum. Once the mechanics are mastered, these exercises develop the student's skills at comprehending realistic problems that describe situations to which the techniques may be applied.

On Your Own . . . Each chapter ends with an exercise entitled "On Your Own. . . ." The intent of this exercise is to give the student some hands-on experience with an application of the statistical concepts introduced in the chapter. In most cases, the student is required to collect, analyze, and interpret data relating to some real application.

A Choice in Level of Coverage of Probability One of the most troublesome aspects of an introductory statistics course is the study of probability. Probability is troublesome for instructors because they must decide on the level of presentation, and it is troublesome for students because they (often) find it difficult at any level. We believe that one cause for these problems is the mixture of probability and counting rules that occurs in most introductory texts. We have included the counting rules in a separate and optional section at the end of the chapter on probability. In addition, all exercises that require the use of counting rules are marked with an asterisk (*) to indicate this. Thus, the instructor can control the level of coverage of probability.

A word should be added about the length of the probability chapter. Although more space is devoted to probability than in many introductory texts, there are three simple explanations for this: more examples, more exercises, and the optional counting rule section mentioned above. We have included the usual die and coin examples to introduce concepts, but we also work many practical examples so that the connection between probability and statistics is clearly made. This same pattern is

followed in the exercise sections. Thus, the 31 examples and 101 exercises account for the length of the chapter. We trust that these will make this troublesome subject easier to learn and to teach.

Where We've Been . . . Where We're Going . . . The first page of each chapter is a "unification" page. Our purpose is to allow the student to see how the chapter fits into the scheme of statistical inference. First, we briefly show how the material presented in previous chapters helps us to achieve our goal (Where we've been). Then, we indicate what the next chapter (or chapters) contributes to the overall objective (Where we're going). This feature allows us to point out that we are constructing the foundation block by block, with each chapter an important component in the structure of statistical inference. Furthermore, this feature provides a series of brief resumés of the material covered as well as glimpses of future topics.

Chapter Table of Contents Preceding each chapter is a chapter table of contents. This enables the student to see the sequence of topics to be covered in the chapter.

An Extensive Coverage of Multiple Regression Analysis and Model Building This topic represents one of the most useful statistical tools for the solution of applied problems. Although an entire text could be devoted to regression modeling, we feel that we have presented a coverage that is understandable, usable, and much more comprehensive than the presentations in other introductory statistics texts. We devote three chapters to discussing the major types of inferences that can be derived from a regression analysis, showing how these results appear in computer printouts and, most important, selecting multiple regression models to be used in an analysis. Thus, the instructor has the choice of a one-chapter coverage of simple regression, a two-chapter treatment of simple and multiple regression, or a complete three-chapter coverage of simple regression, multiple regression, and model building. This extensive coverage of such useful statistical tools will provide added evidence to the student of the relevance of statistics to the solution of applied problems.

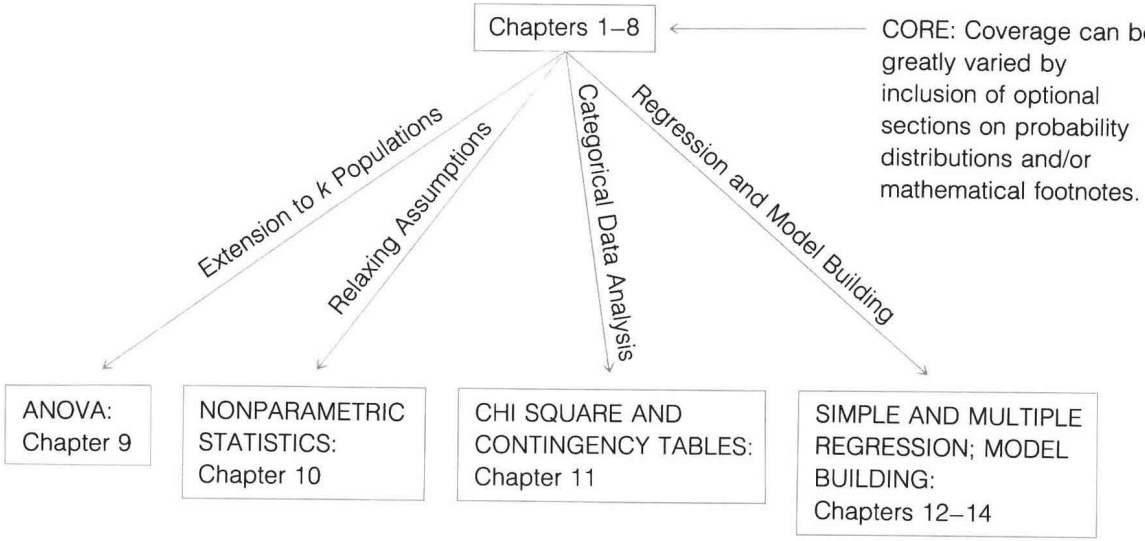
Footnotes Although the text is designed for students with a noncalculus background, footnotes explain the role of calculus in various derivations. Footnotes are also used to inform the student about some of the theory underlying certain results. The footnotes allow additional flexibility in the mathematical and theoretical level at which the material is presented.

Supplementary Materials A study guide, a solutions manual, and a 3,000 item test bank are available.

We have organized the text in what we believe is a logical sequence. For example, rather than place analysis of variance and nonparametric statistics at the end of the

book, we have placed them immediately following the chapters on making inferences about one and two populations. The analysis of variance extends these methods to k populations, and nonparametric statistics gives procedures for making the inferences when the assumptions necessary for parametric methods are in doubt. The three regression and model building chapters, which we consider an essential and unique component, are the final three chapters, because regression represents a greater change in direction than the other inference-making techniques.

Because many introductory courses in statistics are of one-term duration, we show several possible sequences of coverage in the diagram below:



Thanks are due to many individuals who helped in the preparation of this text. Among them are William Beyer, University of Akron; Rudy Gideon, University of Montana; Jean L. Holton, Virginia Commonwealth University; John H. Kellermeier, Northern Illinois University; William G. Koellner, Montclair State University; Diane Lambert, Carnegie–Mellon University; James Lang, Valencia Junior College; Pi-erh Lin, Florida State University; William B. Owen, Central Washington University; Won J. Park, Wright State University; Charles W. Sinclair, Portland State University; Vasanth B. Solomon, Drake University; and Augustin Vukov, University of Toronto. Special thanks to John Dirksey, California State University at Bakersfield. Phyllis Niklas and Susan Reiland have our appreciation and admiration for editing and producing this book. Their work defies explanation; you have to see to believe the care and professionalism with which they work. Finally, we thank the thousands of students at the University of Florida who have helped us to form our ideas about teaching statistics. Their most common complaint seems to be that texts are written for the instructor rather than the student. We hope that this book is an exception.

Case Studies

Third Edition

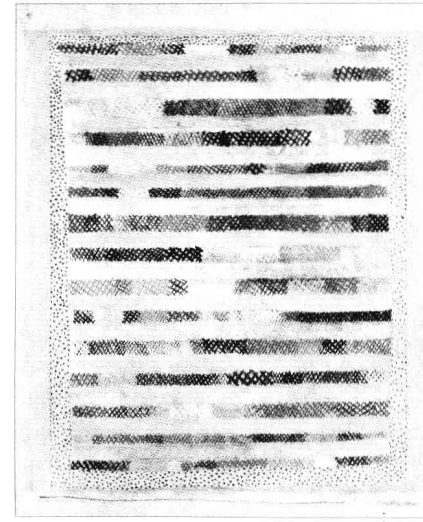
Statistics

Contents

Appendix

CHAPTER 1

What Is Statistics?



Where We're Going . . .

Statistics? Is it a field of study, a group of numbers that summarize the state of our national economy, the performance of a football team, the social conditions in a particular locale, or, as the title of a popular book (Tanur et al., 1978) suggests, "a guide to the unknown"? We will attempt to answer this question in Chapter 1. Throughout the remainder of the text, we will show you how statistics can be used to interpret experimental and sample survey data. Since many jobs in government, industry, medicine, and other fields require this facility, you will see how statistics can be beneficial to you.

Contents

1.1 Statistics: What Is It?

What does statistics mean to you? Does it bring to mind batting averages, Gallup polls, unemployment figures, numerical distortions of facts (lying with statistics!), or simply a college requirement you have to complete? We hope to convince you that statistics is a meaningful, useful science with a broad, almost limitless scope of application to business, government, and the sciences. We also want to show that statistics lie only when they are misapplied. Finally, our objective is to paint a unified picture of statistics to leave you with the impression that your time was well spent studying a subject that will prove useful to you in many ways.

Statistics means "numerical descriptions" to most people. Monthly unemployment figures, the failure rate of a particular type of steel-belted automobile tire, and the proportion of women who favor the Equal Rights Amendment all represent statistical descriptions of large sets of data collected on some phenomenon. Most often the purpose of calculating these numbers goes beyond the description of the set of data. Frequently, the data are regarded as a sample selected from some larger set of data. For example, a sampling of unpaid accounts for a large merchandiser would allow you to calculate an estimate of the average value of unpaid accounts. This estimate could be used as an audit check on the total value of all unpaid accounts held by the merchandiser. So, the applications of statistics can be divided into two broad areas: (1) describing large masses of data and (2) drawing conclusions (making estimates, decisions, predictions, etc.) about some set of data based on sampling. Let us examine some case studies that illustrate applications of statistics.

Case Study 1.1
A Survey: Where "Women's Work" Is Done by Men

The 1980 February/March issue of *Public Opinion* describes the results of a survey of several hundred married men from each of nine countries who responded to the following question:

> In the following list, which household jobs would you say it would be reasonable that the man would often take over from his wife: washing up (doing dishes), changing baby's napkin (diaper), cleaning house, ironing, organizing meal, staying at home with sick child, shopping, none of these?

The graphs in Figure 1.1 provide an effective summary of the thousands of opinions obtained and allow for an easy comparison of attitudes across countries. The area of statistics concerned with the summarization and description of data is called ***descriptive statistics.***

Case Study 1.2
An Experiment: Investigating an Effect of Smoking During Pregnancy

In an article in the *Journal of the American Medical Association,* M. Sexton and J. R. Hebel (1984) report on their research into the effects of maternal smoking on the birth weight of babies. Their experiment randomly assigned 935 pregnant women smokers into two groups: One continued smoking throughout pregnancy and the other, the ***control group,*** received smoking intervention (i.e., assistance to reduce to eliminate smoking). From the measured baby weights of these two groups of women, Sexton and Hebel inferred that "some fetal growth retardation can be overcome by the provision of antismoking assistance to pregnant women."