

# A FIRST COURSE IN STATISTICS

FOURTH EDITION

JAMES T. McCLAVE  
FRANK H. DIETRICH II



0212  
M478  
E.4

9760923

50<sup>25</sup>

FOURTH EDITION

# A First Course in Statistics

---

JAMES T. MCCLAVE

University of Florida

FRANK H. DIETRICH II

Northern Kentucky University



E9760923

DELLEN

an imprint of

MACMILLAN PUBLISHING  
COMPANY

New York

MAXWELL MACMILLAN CANADA

Toronto

MAXWELL MACMILLAN INTERNATIONAL

New York Oxford Singapore Sydney

*On the cover:* Painted wood sculptures by San Francisco artist and designer Robert Hutchinson. Photography by John Jensen.

Copyright © 1992 by Dellen, an imprint of Macmillan, Inc.

Printed in the United States of America

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the Publisher.

Macmillan Publishing Company  
866 Third Avenue, New York, New York 10022

Macmillan Publishing Company is part of the  
Maxwell Communication Group of Companies.

Maxwell Macmillan Canada, Inc.  
1200 Eglinton Avenue East  
Suite 200  
Don Mills, Ontario M3C 3N1

**Library of Congress Cataloging in Publication Data**

McClave, James T.

A first course in statistics / James T. McClave, Frank H.  
Dietrich, II—4th ed.

p. cm.

Includes index.

ISBN 0-02-378561-6

1. Statistics I. Dietrich, Frank H. II. Title.

QA276.12.M3997 1992

519.5—dc20

91-31985

CIP

Printing: 2 3 4 5 6 7 Year: 2 3 4 5 6 7 8

FOURTH EDITION

# A First Course in Statistics

---

# Preface

---

Instructors who teach a single one-semester (one-quarter) introductory statistics course may find that many of the available texts have been written for a two-semester sequence. *A First Course in Statistics* is designed to cope with this problem.

The fourth edition of *A First Course in Statistics* maintains the style and flavor of our other introductory text, *Statistics*, but is greatly reduced in size. This reduction in content was achieved by deleting “optional” topics—that is, topics an instructor might want to cover in a two-semester sequence but not in a one-semester (one-quarter) course. For example, the optional section on counting rules was deleted from the chapter on probability (Chapter 3). This change reduces the level of difficulty for the student while still providing a good introduction to the probabilistic concepts needed for a study of statistical inference. Optional sections on the Poisson, hypergeometric, geometric, exponential, and uniform random variables were deleted, enabling us to combine coverage of the important binomial and normal probability distributions into a single chapter (Chapter 4). Material on the comparison of population means and proportions was split into two chapters, one (Chapter 8) describing the methodology for comparing two or more population means and the other (Chapter 9), the methodology for comparing two or more population proportions. Nonparametric tests are included in brief optional sections in Chapters 7 and 8. The lengthy chapter on the analysis of variance is omitted, and the chi-square analysis of contingency tables is introduced as an optional section at the end of Chapter 9. Simple linear regression analysis is included in Chapter 10, but multiple regression analysis—a topic usually presented in the second semester of a two-semester course—has been omitted.

These changes have produced a text that is suitable for a one-semester introductory course in statistics. In addition, the text allots a modest amount of space to topics that would not usually be covered in an introductory course but might serve as a reference source for the student. Like *Statistics*, the text provides a good introduction to the basic concepts of statistics and to the important concepts and methods of statistical inference.

In addition to making important changes in wording that improve readability of the text, the fourth edition of *A First Course in Statistics* contains the following major changes:

1. **Measurement.** We introduce the concept of *variables* in Chapter 1. We do this by redefining the population to consist of “units” rather than “data,” and then defining a variable to be the characteristic of these units. Data are then obtained by measuring the variable of interest. This will focus more attention on measurement and will make an easier transition to the concept of random variables and their probability distributions.



2. **Types of Data.** In the same vein, in Chapter 2 we have added more on the concept of measuring different types of variables, introducing and defining the four types of data: nominal, ordinal, interval, and ratio.
3. **Estimation and Hypothesis Tests.** The introduction to the classical inferential techniques of estimation and tests of hypotheses had made for a very long Chapter 6 in the third edition, especially with the (optional) material on Type II errors. In the fourth edition we have separated single-sample inference into two chapters: one on estimation (Chapter 6) and the second on tests of hypotheses (Chapter 7).
4. **Exercises.** New exercises have been added throughout, this time concentrating on new and better mechanical exercises in the Learning the Mechanics sections at the beginning of most exercise sets. Some new applied exercises have been added as well, especially accompanying the new material.

The flexibility of past editions is maintained in this edition. Sections that are not prerequisite to succeeding sections and chapters are marked “(Optional).” For example, an instructor who wishes to devote significant time to exploratory data analysis might cover all topics in Chapter 2. In contrast, an instructor who wishes to move rapidly into inferential procedures might omit the optional section on box plots and devote only several lectures to this chapter.

We have maintained the features of this text that we believe make it unique among introductory statistics texts. These features, which assist the student in achieving an overview of statistics and an understanding of its relevance in the sciences, business, and everyday life, are as follows:

1. **Case Studies.** (See the list of case studies on page xv.) 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.
2. **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.
3. **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.
4. **Many Exercises—Labeled by Type.** The text has a large number of exercises illustrating applications in almost all areas of research. (The answers for odd-numbered exercises are included at the end of the text.) However, we believe that many students have trouble learning the mechanics of statistical tech-

niques 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:

- a. **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.
  - b. **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.
5. **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.
  6. **Using the Computer.** We continue to encourage the use of computers in the analysis of real data. A demographic database, consisting of 1,000 observations on 15 variables, has been described in Appendix B and is available on diskette from the publisher. Most chapters include a suggested computer application, **Using the Computer . . .**, which provides one or more computer exercises that utilize the data in Appendix B and enhance the new material covered in the chapter.
  7. **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 résumés of the material covered as well as glimpses of future topics.
  8. **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.
  9. **Supplementary Material.** Solutions manuals, a study guide, a Minitab supplement, an integrated companion software system, a computer-generated test system, and a 1,000-observation demographic database are available.
    - a. **Solutions Manuals** (by Nancy Shafer Boudreau). The solutions manuals present detailed solutions to most exercises in the text. Many points are clarified and expanded to provide maximum insight into and benefits from

each exercise. The student's manual provides solutions to odd-numbered exercises, while the instructor's manual provides solutions to the even-numbered exercises.

- b. Study Guide (by Susan L. Reiland). For each chapter, the study guide includes (1) a brief summary that highlights the concepts and terms introduced in the textbook; (2) section-by-section examples with detailed solutions; and (3) exercises (with answers provided at the end of the study guide) that allow the student to check mastery of the material in each section.
- c. Minitab Supplement (by Ruth K. Meyer and David D. Krueger). The Minitab computer supplement was developed to be used with Minitab Release 7.0, a general-purpose statistical computing system. The supplement, which was written especially for the student with no previous experience with computers, provides step-by-step descriptions of how to use Minitab effectively as an aid in data analysis. Each chapter begins with a list of new commands introduced in the chapter. Brief examples are then given to explain new commands, followed by examples from the text illustrating the new and previously learned commands. Where appropriate, simulation examples are included. Exercises, many of which are drawn from the text, conclude each chapter.

A special feature of the supplement is a chapter describing a survey sampling project. The objectives of the project are to illustrate the evaluation of a questionnaire, provide a review of statistical techniques, and illustrate the use of Minitab for questionnaire evaluation.

- d. DellenStat (by Michael Conlon). DellenStat is an integrated statistics package consisting of a workbook and an IBM PC floppy diskette with software and example sets of data. The system contains a file creation and management facility, a statistics facility, and a presentation facility. The software is menu-driven and has an extensive help facility. It is completely compatible with the text.

The DellenStat workbook describes the operation of the software and uses examples from the text. After an introductory chapter for new computer users, the remaining chapters follow the outline of the text. Additional chapters show how to create new sets of data. Technical appendices cover material for advanced users and programmers.

DellenStat runs on any IBM PC or close compatible with at least 256K of memory and at least one floppy disk drive.

- e. DellenTest. This unique computer-generated random test system is available to instructors without cost. Utilizing an IBM PC computer and a number of commonly used dot-matrix printers, the system will generate an almost unlimited number of quizzes, chapter tests, final examinations, and drill exercises. At the same time, the system produces an answer key and student worksheet with an answer column that exactly matches the column on the answer key.
- f. Database. A demographic data set was assembled based on a systematic random sample of 1,000 U.S. zip codes. Demographic data for each zip code area selected were supplied by CACI, an international demographic



and market information firm. Fifteen demographic measurements (including population, number of households, median age, median household income, variables related to the cost of housing, educational levels, the work force, and purchasing potential indexes based on the Bureau of the Census Consumer Expenditure Surveys) are presented for each zip code area.

Some of the data are referenced in the **Using the Computer** sections. The objectives are to enable the student to analyze real data in a relatively large sample using the computer, and to gain experience using the statistical techniques and concepts on real data.

## Acknowledgments

Thanks are due to many individuals who helped in the preparation of this text. Among them are the reviewers, whose names are listed below. Special thanks to John Dirkse, California State University at Bakersfield. Susan Reiland has our appreciation and admiration for managing the production of this book. Her work defies explanation; you have to see to believe the care and professionalism with which she works. We thank Brenda Dobson for converting our scribbling to a polished manuscript. Finally, we thank the thousands of students 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.

David Atkinson  
Olivet Nazarene University  
William H. Beyer  
University of Akron  
Patricia M. Buchanan  
Pennsylvania State University  
Kathryn Chaloner  
University of Minnesota  
Larry Dion  
California State University—Chico  
Michael J. Doviak  
Old Dominion University  
N. B. Ebrahimi  
Northern Illinois University  
Rudy Gideon  
University of Montana  
David Groggel  
Miami University at Oxford

John E. Groves  
California Polytechnic State  
University at San Luis Obispo  
Shu-ping Hodgson  
Central Michigan University  
Jean L. Holton  
Virginia Commonwealth University  
John H. Kellermeier  
Northern Illinois University  
Timothy J. Killeen  
University of Connecticut  
William G. Koellner  
Montclair State University  
James R. Lackritz  
San Diego State University  
Diane Lambert  
Carnegie–Mellon University  
James Lang  
Valencia Junior College

John J. Lefante, Jr.  
University of South Alabama

Pi-Erh Lin  
Florida State University

R. Bruce Lind  
University of Puget Sound

Linda C. Malone  
University of Central Florida

Allen E. Martin  
California State University at Los  
Angeles

Leslie Matekaitis  
Northern Illinois University

A. Mukherjea  
University of South Florida

Bernard Ostle  
University of Central Florida

William B. Owen  
Central Washington University

Won J. Park  
Wright State University

John J. Peterson  
Smith Kline & French Laboratories

Rita Schillaber  
University of Alberta

James R. Schott  
University of Central Florida

Susan C. Schott  
University of Central Florida

George Schultz  
St. Petersburg Junior College

Carl James Schwarz  
University of Manitoba

Charles W. Sinclair  
Portland State University

Robert K. Smidt  
California Polytechnic State  
University at San Luis Obispo

Vasanth B. Solomon  
Drake University

W. Robert Stephenson  
Iowa State University

Barbara Treadwell  
Western Michigan University

Dan Voss  
Wright State University

Augustin Vukov  
University of Toronto

Theophil J. Worosz  
Metropolitan State  
College of Denver

# Contents

---

Preface ix

## CHAPTER 1 What Is Statistics? 1

- 1.1 Statistics: What Is It? 2
- 1.2 The Elements of Statistics 5
- 1.3 Statistics: Witchcraft or Science? 9
- 1.4 Why Study Statistics? 11

## CHAPTER 2 Methods for Describing Sets of Data 17

- 2.1 Types of Data 18
- 2.2 Graphic Methods for Describing Quantitative Data: Stem and Leaf Displays 23
- 2.3 Graphic Methods for Describing Quantitative Data: Histograms 32
- 2.4 Numerical Measures of Central Tendency 43
- 2.5 Numerical Measures of Variability 52
- 2.6 Interpreting the Standard Deviation 60
- 2.7 Measures of Relative Standing 69
- 2.8 Box Plots: Graphic Descriptions Based on Quartiles (Optional) 75
- 2.9 Distorting the Truth with Descriptive Techniques 86

## CHAPTER 3 Probability 99

- 3.1 Events, Sample Spaces, and Probability 100
- 3.2 Compound Events 116
- 3.3 Complementary Events 120
- 3.4 Conditional Probability 127
- 3.5 Probabilities of Unions and Intersections 136
- 3.6 Probability and Statistics: An Example 149
- 3.7 Random Sampling 150

CHAPTER 4	<b>Random Variables and Probability Distributions</b>	167
	4.1 Two Types of Random Variables	168
	4.2 Probability Distributions for Discrete Random Variables	171
	4.3 The Binomial Distribution	182
	4.4 Probability Distributions for Continuous Random Variables	197
	4.5 The Normal Distribution	198
	4.6 Approximating a Binomial Distribution with a Normal Distribution	214
CHAPTER 5	<b>Sampling Distributions</b>	233
	5.1 What Is a Sampling Distribution?	235
	5.2 Properties of Sampling Distributions: Unbiasedness and Minimum Variance (Optional)	243
	5.3 The Central Limit Theorem	248
	5.4 The Relation Between Sample Size and a Sampling Distribution (Optional)	258
CHAPTER 6	<b>Inferences Based on a Single Sample: Estimation</b>	271
	6.1 Large-Sample Estimation of a Population Mean	272
	6.2 Determining the Sample Size Necessary to Estimate a Population Mean	280
	6.3 Small-Sample Estimation of a Population Mean	285
	6.4 Large-Sample Estimation of a Binomial Probability	293
	6.5 Determining the Sample Size Necessary to Estimate a Binomial Probability	299
CHAPTER 7	<b>Inferences Based on a Single Sample: Tests of Hypotheses</b>	309
	7.1 The Elements of a Test of Hypothesis	310
	7.2 Large-Sample Test of Hypothesis About a Population Mean	317
	7.3 Observed Significance Levels: $p$ -Values	327
	7.4 Small-Sample Test of Hypothesis About a Population Mean	331
	7.5 A Nonparametric Test About a Population Median (Optional)	338
	7.6 Large-Sample Test of Hypothesis About a Binomial Probability	344
	7.7 Calculating Type II Error Probabilities: More About $\beta$ (Optional)	350
CHAPTER 8	<b>Inferences for Means Based on Two Samples</b>	367
	8.1 Large-Sample Inferences About the Difference Between Two Population Means: Independent Sampling	368

- 8.2 Small-Sample Inferences About the Difference Between Two Population Means: Independent Sampling 379
- 8.3 Inferences About the Difference Between Two Population Means: Paired Difference Experiments 390
- 8.4 Determining the Sample Size 404
- 8.5 A Nonparametric Test for Comparing Two Populations: Independent Sampling (Optional) 408
- 8.6 A Nonparametric Test for Comparing Two Populations: Paired Difference Experiments (Optional) 417

## CHAPTER 9 Comparing Population Proportions 437

- 9.1 Inferences About the Difference Between Two Population Proportions: Independent Binomial Experiments 438
- 9.2 Determining the Sample Size 448
- 9.3 Comparing Population Proportions from a Multinomial Experiment (Optional) 450
- 9.4 Contingency Table Analysis (Optional) 461

## CHAPTER 10 Simple Linear Regression 483

- 10.1 Probabilistic Models 484
- 10.2 Fitting the Model: Least Squares Approach 488
- 10.3 Model Assumptions 499
- 10.4 An Estimator of  $\sigma^2$  501
- 10.5 Assessing the Usefulness of the Model: Making Inferences About the Slope  $\beta_1$  505
- 10.6 The Coefficient of Correlation 512
- 10.7 The Coefficient of Determination 515
- 10.8 Using the Model for Estimation and Prediction 523
- 10.9 Simple Linear Regression: An Example 533
- 10.10 Using the Computer for Simple Linear Regression 537

## APPENDIX A Tables 553

- Table I Random Numbers 555
- Table II Binomial Probabilities 558
- Table III Normal Curve Areas 562
- Table IV Critical Values of  $t$  563
- Table V Critical Values of  $T_L$  and  $T_U$  for the Wilcoxon Rank Sum Test: Independent Samples 564



Table VI Critical Values of  $T_0$  in the Wilcoxon Paired Difference  
Signed Rank Test 565

Table VII Critical Values of  $\chi^2$  566

APPENDIX **B** **Demographic Data Set** 569

**Answers to Selected Exercises** 573

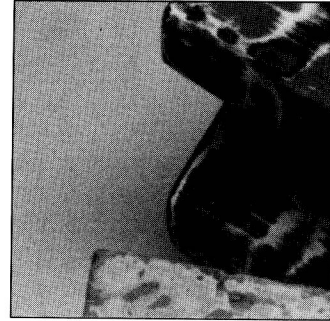
**Index** 581

# Case Studies

---

- 1.1 A Survey: Where “Women’s Work” Is Done by Men 2
- 1.2 An Experiment: Investigating an Effect of Smoking During Pregnancy 4
- 1.3 Does Judicial Action Affect the Probability of Conviction? 4
- 1.4 Taste Preference for Beer: Brand Image or Physical Characteristics of the Beer? 5
- 2.1 Mercury Poisoning and the Dental Profession 37
- 2.2 Hotels: A Rational Method for Overbooking 45
- 2.3 Statistics and Air Quality Standards 72
- 2.4 *Children Out of School in America*: Making an Ugly Picture Look Worse 89
- 3.1 Bloom County Probabilities 104
- 3.2 Comparing Subjective Probability Assessments with Relative Frequencies 109
- 3.3 Purchase Patterns and the Conditional Probability of Purchasing 130
- 3.4 The 1970 Draft Lottery 154
- 4.1 A Restaurant Chain Fights Sales Tax Claim 177
- 4.2 The Space Shuttle *Challenger*: Catastrophe in Space 187
- 4.3 A Survey of Children’s Political Knowledge 192
- 4.4 Grading on the Curve 209
- 6.1 Dancing to the Customer’s Tune: The Need to Assess Customer Preferences 277
- 7.1 Statistics Is Murder! 315
- 7.2 Hypothesis Tests in Computer Security Systems 323
- 8.1 Productivity and Mobility 374
- 8.2 Matched Pairing in Studying the Mentally Retarded 397
- 9.1 Hotel Room Interviews—Anxiety and Suspicion 442
- 9.2 Investigating Response Bias in a Diary Survey 456
- 9.3 Evaluating a New Method for Treating Cancer 469
- 10.1 Predicting the United States Crime Index 518

# What Is Statistics?



## CONTENTS

- 1.1 Statistics: What Is It?
- 1.2 The Elements of Statistics
- 1.3 Statistics: Witchcraft or Science?
- 1.4 Why Study Statistics?
- Case Study 1.1 A Survey: Where “Women’s Work” Is Done by Men
- Case Study 1.2 An Experiment: Investigating an Effect of Smoking During Pregnancy
- Case Study 1.3 Does Judicial Action Affect the Probability of Conviction?
- Case Study 1.4 Taste Preference for Beer: Brand Image or Physical Characteristics of the Beer?

## WHERE WE’RE GOING...

Statistics? Is it a field of study, a group of numbers that summarizes 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., 1989) suggests, “a guide to the unknown”? We will see in Chapter 1 that each of these descriptions has some applicability in understanding statistics. We will see that **descriptive statistics** focuses on developing numerical summaries that describe some phenomenon, whereas **inferential statistics** uses these numerical summaries to assist in making decisions. The primary theme of this text is inferential statistics. Thus, we concentrate on showing you how statistics can be used to interpret and use data to make decisions. Since many jobs in industry, government, medicine, and other fields require this facility, we show you how statistics can be beneficial to you.

## 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. 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 whose characteristics we may wish to estimate. 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. So, the applications of statistics can be divided into two broad areas: *descriptive* and *inferential* statistics.



**Descriptive statistics** utilizes numerical and graphical methods to look for patterns, to summarize, and to present the information in a set of data.

**Inferential statistics** utilizes sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data.

Although both descriptive and inferential statistics are discussed in the following chapters, the primary theme of the text is *inference*. 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?