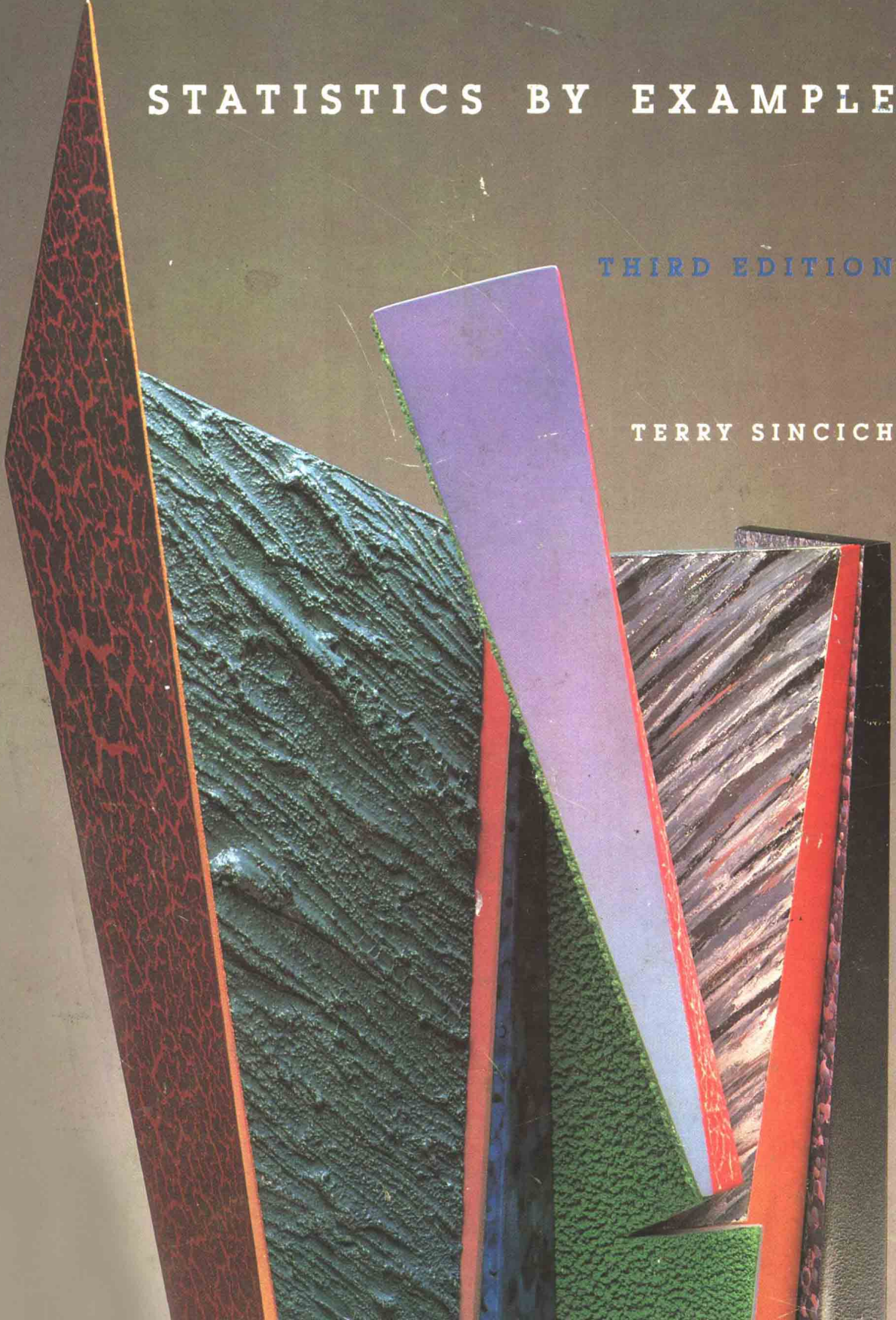


STATISTICS BY EXAMPLE

THIRD EDITION

TERRY SINCICH



THIRD EDITION

Statistics by Example

Terry Sincich

University of Florida

Dellen Publishing Company

San Francisco

Collier Macmillan Publishers

London

divisions of Macmillan, Inc.

On the cover: "Back Talk" by Stuart Lehrman, 1985; polychrome wood, 34" × 15" × 12". Stuart Lehrman combines sculpture with painting in works comprised of vibrantly colored and richly textured three-dimensional planes. Lehrman's work is included in many public and private collections in the United States. Lehrman is represented by Acme Art in San Francisco.

Copyright 1987, 1985, 1982 by Dellen Publishing Company,
a division 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.

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

Sincich, Terry.
Statistics by example.

Includes index.

1. Statistics. I. Title.
QA276.12.S56 1987 519.5 86-13569

Printing 1 2 3 4 5 6 7 8 9 Year 7 8 9 0 1

ISBN 0-02-411031-0

PREFACE

This introductory college statistics text is designed for students who have only a high school background in mathematics as a prerequisite. It differs from most other texts in two ways:

- 1. Real data sets** Explanations of basic statistical concepts and methodology are based on and motivated by the use of real data sets.
- 2. Teaching by example** Concepts and statistical methods are explained in examples. These examples arise as questions posed about the data sets.

We think that this practical orientation helps the student to relate statistics to real-life problems and, hopefully, will develop a pattern of thought that will persist after the student enters the business world.

The text contains four data sets; the first two are heavily used as instructional vehicles. These data sets are:

Appendix A The set of actual starting salaries, majors, and colleges of 1,262 University of Florida graduates during the period from December 1981 to August 1983.

Appendix B The starting salaries (extracted from Appendix A) of bachelor's degree graduates of the Colleges of Business Administration, Education, Engineering, Liberal Arts, and Sciences.

Appendix C Supermarket customer checkout times for mechanical and automated checkers.

Appendix D Length, weight, and DDT measurements for various species of fish collected from the Tennessee River, Alabama, and its creek tributaries.

Although all of the data sets are used to develop the notion of a population and a sample, the starting salaries of college graduates (Appendices A and B) are used to demonstrate the need for data description, to develop the notion of a sampling distribution, and to motivate the inferential methods commonly studied in an introductory statistics course.

In addition to teaching via data sets and by example, this edition retains the following features of the first two editions:

- 1. Case studies** At least three case studies that detail specific current events are used at the end of each chapter to pose questions for the student. These case studies, extracted from news articles and journals, demonstrate to the student the relevance of statistics to the solution of current practical problems. Answers to most case studies are included at the end of the text. (Refer to the Table of Contents for a list of the case studies.)
- 2. Examples** The text, as its name implies, employs the "teaching by example" method. Each section contains several worked examples to demonstrate how to solve various types of statistical problems encountered in the real world.

- 3. Key concepts highlighted** Throughout the text, key concepts are highlighted through the use of screened boxes.
 - a. Definitions** are boxed.
 - b. Steps** for constructing bar graphs, performing statistical calculations, and conducting statistical tests are listed and boxed for each procedure.
 - c. Key words**, which must be added to a student's vocabulary, are listed and boxed at the end of each chapter.
 - d. Warnings**, indicating situations where a student might misuse a statistical technique, are presented in boxed form. The student is directed to specific alternative methods.
- 4. Probability and blackjack** The basic concepts of probability and their relation to statistical inference are presented in an easy-to-understand manner and are developed around the game of blackjack. Problem solving for the sake of problem solving is avoided.
- 5. Many exercises** Since most students learn best by doing, the text contains a large number (over 1,000) of exercises. The answers for most are included at the end of the text. Each chapter contains exercises at the end of each section and a set of supplementary exercises at the chapter's end. The exercises are of two types:
 - a. Learning the Mechanics** These exercises are intended to be straightforward applications of the new concepts presented in the section. 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 definition.
 - b. Applying the Concepts** The mechanical exercises are followed by realistic exercises that allow the student to see applications of statistics to the solutions of a variety of real-world problems. Many of these exercises contain data extracted from newspaper articles, magazines, and journals. Once the mechanics are mastered, these exercises develop students' skills at comprehending realistic problems that describe situations to which the techniques may be applied.
- 6. Computer printouts** The use of computer program packages is introduced in the presentation of the analysis of variance (Chapter 11), simple linear regression analysis (Chapter 12), multiple regression analysis (Chapter 13), and contingency table analysis (Chapter 14). The computer printouts for four different program packages, Minitab, SAS, SPSS^x, and BMDP, are presented for the analyses of identical sets of data.
- 7. Integrating the data sets and the computer** The data sets can be entered into computer storage and can be accessed by students for sampling and statistical inference. The student (or instructor) can then access the data sets for the demonstration of statistical concepts and for realistic statistical exercises. For example, the data sets can be used by the instructor to illustrate the concept of a sampling distribution and the concepts of estimation and tests of hypotheses.

Although the scope of coverage remains the same, the third edition contains several substantial changes, additions, and enhancements:

- 1. Case study section in all chapters** Each chapter now contains a case study section, in addition to the case studies presented at the end of the chapter. These new sections describe actual problems encountered in various fields of study and apply the statistical techniques learned in the chapter to solve the problems. Thus, each case study section gives an excellent example of the use of statistics in the real world.
- 2. Chapter 7: Sampling and Sampling Distributions** A new section on the problems of nonresponse and invalid responses in survey sampling (Section 7.7) has been added.
- 3. Chapter 11: Analysis of Variance** This chapter has been expanded to include randomized block designs. Three new sections have been added. The computing formulas for conducting an ANOVA with a randomized block design are given in Section 11.5, the formulas for confidence intervals and tests for differences between means in Section 11.6, and an example of a computer printout in Section 11.7.
- 4. Chapter 15: Nonparametric Statistics** New sections on the Kruskal–Wallis H test for analyzing independent samples designs (Section 15.4) and the Friedman F_r test for analyzing randomized block designs (Section 15.5) have been added.
- 5. More exercises with real data** Many new “real-world” exercises have been added to each chapter. These exercises, like the case studies, are extracted from news articles, magazines, and professional journals to give students the opportunity to apply their knowledge of statistics to current practical problems.

Numerous less obvious changes in details have been made throughout the text in response to suggestions by current users of the first two editions.

The text is also accompanied by the following supplementary material:

- 1. Solutions manual** A student’s exercise solutions manual presents the solutions for half the exercises contained in the text.
- 2. Computer supplement** Easy-to-follow instructions on how to enter data into the computer and how to run most of the statistical procedures presented in the text are provided in a computer supplement. Each chapter in the supplement includes sample programs for four popular statistical program packages—SAS, SPSS^X, Minitab, and BMDP.
- 3. Data sets available on diskette or tape** The data in the appendices are available on either an IBM PC diskette or a 7-track, nonlabelled magnetic computer tape.

I wish to acknowledge the many individuals who provided their invaluable assistance during the preparation of the original text and this revision. Their efforts are much appreciated. In particular, I thank the following for providing helpful suggestions and advice on the writing of the manuscript: Randal S. Beck, Millikin University;

Noni Bohonak, Dickinson College; David E. Booth, Millikin University; John S. Bowdidge, Southwest Missouri State University; Phillip G. Buckhiester, North Georgia College; Michael C. Burke, College of San Mateo; John Cameron, Rockhurst College; DeAnn Christianson, University of the Pacific; Jerry I. Goldman, DePaul University; William Hemmer, San Diego Mesa College; Geoffrey B. Holmewood, Hudson Valley Community College; Thomas B. Laase, University of Southern Colorado; James Lang, Valencia Junior College; Paul Lawrisuk, Moraine Valley Community College; James T. McClave, University of Florida; William Mendenhall, University of Florida; Franklin D. Rich, Sam Houston State University; and John B. Rushton, Metropolitan State College. Susan Reiland deserves special recognition for her excellent line-by-line reviews during the writing of the original manuscript and for managing the production of the second and third editions. I am very grateful to the following for providing the data sets and accompanying background information: Maurice Mayberry (Director, Career Resources Center, University of Florida), Jim Sullivan (Water and Air Research), Info Tech, and Venus Wong (who patiently spent long hours at the supermarket recording customer checkout times). Thanks are also due Faith Sincich for preparing answers to the exercises. I also wish to thank my typists, Brenda Dobson and Carol Springer, who did a remarkable job of converting my hand-scribbled notes into immaculate type.

Finally, I owe very special thanks to William Mendenhall and James T. McClave who, together, suggested the concept of this text and then provided me with the opportunity to write the manuscript. Without their guidance and encouragement this text and its subsequent revisions would never have been completed.

THIRD EDITION

Statistics by Example

CONTENTS

Chapter 1	Introduction	1
	1.1 What Is Statistics?	2
	1.2 How Can Statistics Be of Value in Your Field of Study?	6
	1.3 Case Study: Obedience to Authority—The Shocking Truth	8
	1.4 Summary	10
	Case Study 1.1 Contamination of Fish in the Tennessee River	14
	Case Study 1.2 Cruising: How Foresters Estimate Timber Weights	15
	Case Study 1.3 Auditing Hospital Inventory	16
<hr/>		
Chapter 2	Graphical Methods for Describing Data Sets	19
	2.1 The Objective of Data Description	20
	2.2 Types of Data	20
	2.3 Graphical Descriptions of Qualitative Data	24
	2.4 Graphical Descriptions of Quantitative Data: Stem and Leaf Displays	34
	2.5 Case Study: A Bad Moon Rising	41
	2.6 Graphical Descriptions of Quantitative Data: Relative Frequency Distributions	42
	2.7 Summary	53
	Case Study 2.1 The Most Serious Health Problem Facing Women	65
	Case Study 2.2 College-Age Entrepreneurs—A Trend of the 1980's?	66
	Case Study 2.3 Carbon Monoxide: A Potential Hazard for Cigarette Smokers	67
<hr/>		
Chapter 3	Numerical Methods for Describing Quantitative Data	73
	3.1 Why We Need Numerical Descriptive Measures	74
	3.2 Types of Numerical Descriptive Measures	74
	3.3 Summation Notation	75
	3.4 Measures of Central Tendency	77
	3.5 Measures of Data Variation	84
	3.6 Case Study: How Much Weight Can You Lift—and Still Avoid Serious Injury?	97

3.7	Measures of Relative Standing	100
3.8	Quartiles and the Interquartile Range (Optional)	104
3.9	Looking for Outliers: Box Plots (Optional)	107
3.10	Numerical Descriptive Measures for Populations	111
3.11	Calculating a Mean and Standard Deviation from Grouped Data (Optional)	112
3.12	Summary	116
	Case Study 3.1 Where's the Beef?	135
	Case Study 3.2 Consumer Complaints: Due to Chance or Specific Causes?	137
	Case Study 3.3 Estimating the Mean Weight of Short-Leaf Pine Trees—Grouped Data Method (Optional)	138

Chapter 4	Probability	141
4.1	Blackjack, Investing, and Statistics	142
4.2	The Game of Blackjack	142
4.3	Experiments and Events	143
4.4	The Probability of an Event	148
4.5	The Additive Probability Rule for Mutually Exclusive Events	152
4.6	Random Sampling and the Probability of Blackjack	160
4.7	Conditional Probability and Independence	166
4.8	The Additive and Multiplicative Laws of Probability (Optional)	173
4.9	Case Study: The Illegal Numbers Game	180
4.10	Summary	181
	Case Study 4.1 The Iranian Hostage Rescue Mission	195
	Case Study 4.2 Dr. Crypton's Logic	196
	Case Study 4.3 Oil Leases: Striking It Rich with a \$10 Stake	197
	Case Study 4.4 The Blackjack Victory of the Seven Samurai	198

Chapter 5	Opinion Polls and the Binomial Probability Distribution	201
5.1	Random Variables	202
5.2	Probability Models for Populations	203
5.3	The Binomial Probability Distribution	205
5.4	Tables of the Binomial Probability Distribution	213
5.5	Cumulative Binomial Probability Tables	214
5.6	The Mean and Standard Deviation for a Binomial Probability Distribution	219
5.7	Case Study: The Probability of a Successful Helicopter Flight during the Iran Rescue Mission	228

5.8	Summary	229
Case Study 5.1	Signalling the Rise or Fall of the Stock Market: Corporate Insider Theory	233
Case Study 5.2	Expected Winnings in Junk-Mail Contests	234
Case Study 5.3	Expected Gains in the Oil Lease Lottery	234

Chapter 6 The Normal Distribution 237

6.1	Probability Models for Continuous Random Variables	238
6.2	The Normal Distribution	239
6.3	Case Study: A Random Walk down Wall Street	256
6.4	Summary	262
Case Study 6.1	Comparing Reality with the Normal Curve	266
Case Study 6.2	Interpreting Those Wonderful EPA Mileage Estimates	267
Case Study 6.3	Break-Even Analysis—When to Market a New Product	268

Chapter 7 Sampling and Sampling Distributions 271

7.1	Why the Method of Sampling Is Important	272
7.2	Obtaining a Random Sample	274
7.3	Sampling Distributions	279
7.4	The Sampling Distribution of \bar{x} ; the Central Limit Theorem	285
7.5	Case Study: The Role of Patient Waiting Time	299
7.6	Other Types of Samples: Stratified, Cluster, and Systematic Samples	301
7.7	Problems of Nonresponse and Invalid Responses	303
7.8	Summary	304
Case Study 7.1	Pollsters Blast ABC-TV's Survey	312
Case Study 7.2	A Decision Problem for Financial Managers: When to Investigate Cost Variances?	313
Case Study 7.3	Estimating the Total Weight of Short-Leaf Pine Trees (Optional)	315

Chapter 8 Estimation of Means and Proportions 319

8.1	Introduction	320
8.2	Estimation of a Population Mean: Large-Sample Case	321
8.3	Estimation of a Population Mean: Small-Sample Case	332
8.4	Estimation of a Population Proportion: Large-Sample Case	339
8.5	Estimation of the Difference between Two Population Means: Large-Sample Case	345

8.6	Estimation of the Difference between Two Population Means: Small-Sample Case	353
8.7	Estimation of the Difference between Two Population Means: Matched Pairs	361
8.8	Estimation of the Difference between Two Population Proportions: Large-Sample Case	368
8.9	Choosing the Sample Size	375
8.10	Case Study: The Crazy Daisy Shave	381
8.11	Summary	382
	Case Study 8.1 Consumer Attitudes toward Automated Supermarket Checkers	393
	Case Study 8.2 Sample Surveys: The Foot-in-the-Door and the Door-in-the-Face Approaches	394
	Case Study 8.3 An I.Q. Comparison of Identical Twins Reared Apart	396

Chapter 9**Collecting Evidence to Support a Theory: General
Concepts of Hypothesis Testing****399**

9.1	Introduction	400
9.2	Formulation of Hypotheses	400
9.3	Conclusions and Consequences for a Hypothesis Test	405
9.4	Test Statistics and Rejection Regions	408
9.5	Case Study: Schlitz versus Budweiser—Mug to Mug	418
9.6	Summary	420
	Case Study 9.1 Lead Poisoning and the Tooth Fairy	423
	Case Study 9.2 Drug Screening: A Statistical Decision Problem	424
	Case Study 9.3 Consumer Product Safety Warnings—How Effective Are They?	426

Chapter 10**Hypothesis Testing: Applications****429**

10.1	Introduction	430
10.2	Hypothesis Test about a Population Mean: Large-Sample Case	430
10.3	Hypothesis Test about a Population Mean: Small-Sample Case	438
10.4	Hypothesis Test about a Population Proportion: Large-Sample Case	444
10.5	Hypothesis Test about the Difference between Two Population Means: Large-Sample Case	450
10.6	Case Study: The Downfall of the Roman Empire	457
10.7	Hypothesis Test about the Difference between Two Population Means: Small-Sample Case	459

10.8	Hypothesis Test about the Difference between Two Population Means: Matched Pairs	467
10.9	Hypothesis Test about the Difference between Two Proportions: Large-Sample Case	473
10.10	Reporting Test Results: p -Values	480
10.11	Diagnosing a Hypothesis Test: Determining the Parameter of Interest	485
10.12	Summary	487
	Case Study 10.1 The Invisible Birds	495
	Case Study 10.2 Stress in Nursing Schools: Myth or Mystery?	497
	Case Study 10.3 The $M^*A^*S^*H$ Generation's View of the Business World	498

Chapter 11	Comparing More Than Two Population Means: An Analysis of Variance	501
11.1	Introduction	502
11.2	One-Way Analysis of Variance and the F Statistic	503
11.3	Confidence Intervals and Tests for Means: Independent Sampling Design	520
11.4	Analysis of Variance for an Independent Sampling Design: A Computer Printout	526
11.5	A Two-Way Analysis of Variance for a Randomized Block Design	530
11.6	Confidence Intervals and Tests for the Difference between Means: Randomized Block Design	541
11.7	Analysis of Variance for a Randomized Block Design: A Computer Printout	545
11.8	Case Study: Comparing the Mean Starting Salaries of College Graduates	547
11.9	Assumptions: When the Analysis of Variance F Test Is Appropriate	550
11.10	Summary	550
	Case Study 11.1 Moral Development of Teenagers	559
	Case Study 11.2 A Comparison of the Shopping Habits of Four Types of Grocery Shoppers	561
	Case Study 11.3 Identifying Management Potential: The In-Tray Exercise	563
	Case Study 11.4 A Child's-Eye View of the President	565

Chapter 12	Simple Linear Regression and Correlation	569
12.1	Introduction: Bivariate Relationships	570
12.2	Simple Linear Correlation	570
12.3	Straight-Line Probabilistic Models	584
12.4	How to Fit the Model: The Least Squares Approach	598

12.5	Estimating σ^2	599
12.6	Making Inferences about the Slope β_1	601
12.7	How Well Does the Least Squares Line Fit the Data?	610
12.8	Using the Model for Estimation and Prediction	613
12.9	Simple Linear Regression: A Computer Printout	620
12.10	Case Study: Deducing the Federal Deficit	624
12.11	Assumptions Required for a Linear Regression Analysis	628
12.12	Summary	629
	Case Study 12.1 Television, Children, and Drugs	636
	Case Study 12.2 Evaluating the Protein Efficiency of a New Food Product	638
	Case Study 12.3 The Exploding Cost of an Industrial Sales Call	638
	Case Study 12.4 Mental Imagery and the Third Dimension	640

Chapter 13 Multiple Regression 643

13.1	Introduction: Linear Statistical Models and Multiple Regression Analysis	644
13.2	The General Linear Model	645
13.3	Model Assumptions	647
13.4	Fitting the Model	648
13.5	Estimating and Testing Hypotheses about the β Parameters	650
13.6	Measuring How Well the Model Fits the Data	653
13.7	Testing Whether the Model Is Useful for Predicting y	655
13.8	Using the Model for Estimation and Prediction	657
13.9	Other Computer Printouts	660
13.10	Other Linear Models	667
13.11	Case Study: It Never Rains in California	682
13.12	Summary	687
	Case Study 13.1 Tree Clones and Their Water Balance	698
	Case Study 13.2 Business Economists: Overworked and Underpaid?	699
	Case Study 13.3 The Stock Market Effects of Airline Deregulation	702
	Case Study 13.4 Regression and the Law	703

Chapter 14 Categorical Data and the Chi-Square Distribution 709

14.1	An Example of Categorical Data: The Case of the Herring Gulls	710
14.2	The Chi-Square Statistic	711
14.3	The Analysis of the Herring Gull Data	721
14.4	Assumptions: Situations for Which the Chi-Square Test Is Appropriate	735

14.5	Computer Printouts for the Analysis of Categorical Data	736
14.6	Case Study: Where Grocery Shoppers Shop—High-, Medium-, or Low-Priced Stores	738
14.7	Summary	739
	Case Study 14.1 The Relationship between Neighborhood Design and Crime	749
	Case Study 14.2 Testing the Effectiveness of an Antiviral Drug	751
	Case Study 14.3 Tomorrow's Executives—How They View Job Politics	751

Chapter 15 Nonparametric Statistics 755

15.1	Introduction	756
15.2	Comparing Two Populations Based on Independent Random Samples: The Mann–Whitney U Test	756
15.3	Comparing Two Populations Based on Matched Pairs: The Wilcoxon Signed Ranks Test	766
15.4	Comparing Populations Using an Independent Sampling Design: The Kruskal–Wallis H Test	774
15.5	Comparing Populations Using a Randomized Block Design: The Friedman F_r Test	780
15.6	Rank Correlation: Spearman's r_s Statistic	786
15.7	Case Study: The Art of Forecasting Interest Rates	795
15.8	Summary	799
	Case Study 15.1 Population Growth and the Cost of Land	811
	Case Study 15.2 Identifying Management Potential: The In-Tray Exercise	811
	Case Study 15.3 The Public's Perception of Risk: Real or Imagined?	813

Appendix A	Data Set:	
	Starting Salaries of 1,262 University of Florida Graduates, December 1981 to August 1983	817

Appendix B	Data Set:	
	Starting Salaries of University of Florida Bachelor's Degree Graduates in Five Colleges	831

Appendix C	Data Set:	
	Supermarket Customer Checkout Times for Mechanical and Automated Checking	837

Appendix D**Data Set:**

Results of DDT Analyses on Fish Samples Collected Summer 1980, Tennessee River, Alabama	841
--	-----

Appendix E**Statistical Tables****845**

Table 1	Binomial Probabilities	846
Table 2	Cumulative Binomial Probabilities	849
Table 3	Normal Curve Areas	852
Table 4	Critical Values for Student's t	853
Table 5	Critical Values for the F Statistic: $F_{.10}$	854
Table 6	Critical Values for the F Statistic: $F_{.05}$	856
Table 7	Critical Values for the F Statistic: $F_{.025}$	858
Table 8	Critical Values for the F Statistic: $F_{.01}$	860
Table 9	Critical Values for the χ^2 Statistic	862
Table 10	Random Numbers	864
Table 11	Critical Values of the Sample Coefficient of Correlation, r	867
Table 12	Distribution Function of U	868
Table 13	Critical Values of T_0 in the Wilcoxon Paired Difference Signed Ranks Test	873
Table 14	Critical Values of Spearman's Rank Correlation Coefficient	874
	Answers to Selected Exercises and Case Studies	875
	Index	901

CHAPTER 1

Introduction

In one of the most famous experiments in psychology, Dr. Stanley Milgram (1974) studied the factors that determine the extent to which people obey authority—even if that authority is pushing them to do something they are against. In Milgram's study, subjects playing the role of "teachers" were told to give electric shocks (up to 450 volts) to "learners" who answered questions incorrectly. The object of the experiment was to see how many volts a subject would be willing to give before refusing to comply with the request.

Imagine the massive and impossible task of testing everyone in the country to see how they would respond to the shock experiment. And if we could test everyone, how would we determine what constitutes a "normal" reaction? This text describes the use of sampling and the science of statistics to solve many practical problems encountered in the real world. The case study in Section 1.3 discusses how Milgram measured people's obedience to authority and provides us with an opportunity to apply statistics to real-life data.



Contents

- 1.1** What Is Statistics?
- 1.2** How Can Statistics Be of Value in Your Field of Study?
- 1.3** Case Study: Obedience to Authority—The Shocking Truth
- 1.4** Summary

Case Study 1.1 Contamination of Fish in the Tennessee River

Case Study 1.2 Cruising: How Foresters Estimate Timber Weights

Case Study 1.3 Auditing Hospital Inventory