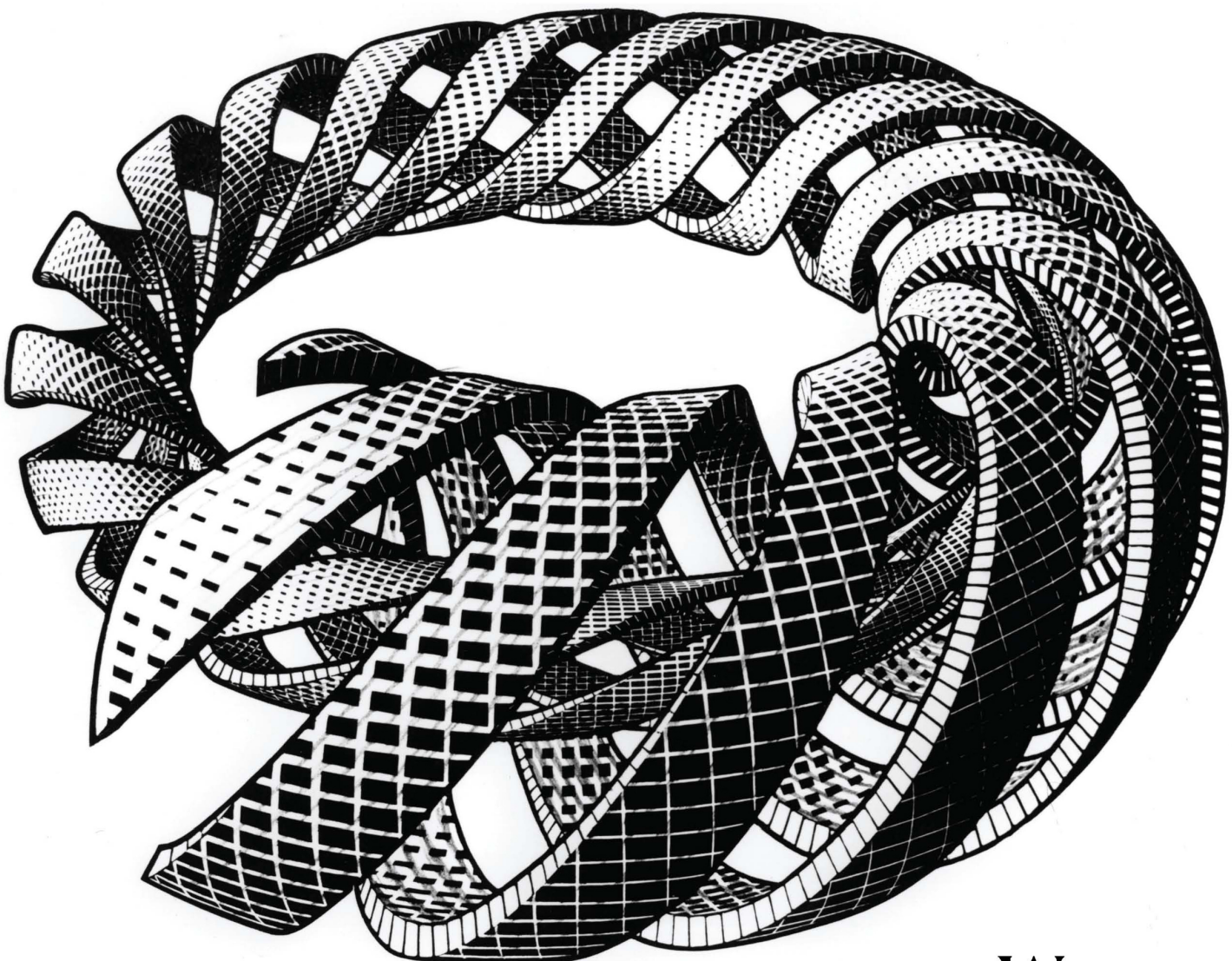
ELEVENTH EDITION

STATISTICS

ROBERT S. **WITTE**

JOHN S. **WITTE**



WILEY

STATISTICS

Eleventh Edition

Robert S. Witte
Emeritus, San Jose State University

John S. Witte
University of California, San Francisco

WILEY

VP AND EDITORIAL DIRECTOR	George Hoffman
EDITORIAL DIRECTOR	Veronica Visentin
EDITORIAL ASSISTANT	Ethan Lipson
EDITORIAL MANAGER	Gladys Soto
CONTENT MANAGEMENT DIRECTOR	Lisa Wojcik
CONTENT MANAGER	Nichole Urban
SENIOR CONTENT SPECIALIST	Nicole Repasky
PRODUCTION EDITOR	Abidha Sulaiman
COVER PHOTO CREDIT	M.C. Escher's Spirals © The M.C. Escher Company - The Netherlands

This book was set in 10/11 Times LT Std by SPi Global and printed and bound by Lightning Source Inc. The cover was printed by Lightning Source Inc.

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

Copyright © 2017, 2010, 2007 John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923 (Web site: www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008, or online at: www.wiley.com/go/permissions.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at: www.wiley.com/go/returnlabel. If you have chosen to adopt this textbook for use in your course, please accept this book as your complimentary desk copy. Outside of the United States, please contact your local sales representative.

ISBN: 978-1-119-25451-5(PBK)
ISBN: 978-1-119-25445-4(EVALC)

Library of Congress Cataloging-in-Publication Data

Names: Witte, Robert S. | Witte, John S.
Title: Statistics / Robert S. Witte, Emeritus, San Jose State University, John S. Witte, University of California, San Francisco.
Description: Eleventh edition. | Hoboken, NJ: John Wiley & Sons, Inc., [2017] | Includes index.
Identifiers: LCCN 2016036766 (print) | LCCN 2016038418 (ebook) | ISBN 9781119254515 (pbk.) | ISBN 9781119299165 (epub)
Subjects: LCSH: Statistics.
Classification: LCC QA276.12 .W57 2017 (print) | LCC QA276.12 (ebook) | DDC 519.5—dc23
LC record available at https://lccn.loc.gov/2016036766

The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

To Doris

Preface

TO THE READER

Students often approach statistics with great apprehension. For many, it is a required course to be taken only under the most favorable circumstances, such as during a quarter or semester when carrying a light course load; for others, it is as distasteful as a visit to a credit counselor—to be postponed as long as possible, with the vague hope that mounting debts might miraculously disappear. Much of this apprehension doubtless rests on the widespread fear of mathematics and mathematically related areas.

This book is written to help you overcome any fear about statistics. Unnecessary quantitative considerations have been eliminated. When not obscured by mathematical treatments better reserved for more advanced books, some of the beauty of statistics, as well as its everyday usefulness, becomes more apparent.

You could go through life quite successfully without ever learning statistics. Having learned some statistics, however, you will be less likely to flinch and change the topic when numbers enter a discussion; you will be more skeptical of conclusions based on loose or erroneous interpretations of sets of numbers; you might even be more inclined to initiate a statistical analysis of some problem within your special area of interest.

TO THE INSTRUCTOR

Largely because they panic at the prospect of any math beyond long division, many students view the introductory statistics class as cruel and unjust punishment. A half-dozen years of experimentation, first with assorted handouts and then with an extensive set of lecture notes distributed as a second text, convinced us that a book could be written for these students. Representing the culmination of this effort, the present book provides a simple overview of descriptive and inferential statistics for mathematically unsophisticated students in the behavioral sciences, social sciences, health sciences, and education.

PEDAGOGICAL FEATURES

- Basic concepts and procedures are explained in plain English, and a special effort has been made to clarify such perennially mystifying topics as the standard deviation, normal curve applications, hypothesis tests, degrees of freedom, and analysis of variance. For example, the standard deviation is more than a formula; it roughly reflects the average amount by which individual observations deviate from their mean.
- Unnecessary math, computational busy work, and subtle technical distinctions are avoided without sacrificing either accuracy or realism. Small batches of data define most computational tasks. Single examples permeate entire chapters or even several related chapters, serving as handy frames of reference for new concepts and procedures.

- Each chapter begins with a preview and ends with a summary, lists of important terms and key equations, and review questions.
- Key statements appear in bold type, and step-by-step summaries of important procedures, such as solving normal curve problems, appear in boxes.
- Important definitions and reminders about key points appear in page margins.
- Scattered throughout the book are examples of computer outputs for three of the most prevalent programs: Minitab, SPSS, and SAS. These outputs can be either ignored or expanded without disrupting the continuity of the text.
- Questions are introduced within chapters, often section by section, as Progress Checks. They are designed to minimize the cumulative confusion reported by many students for some chapters and by some students for most chapters. Each chapter ends with Review Questions.
- Questions have been selected to appeal to student interests: for example, probability calculations, based on design flaws, that re-create the chillingly high likelihood of the *Challenger* shuttle catastrophe (8.18, page 165); a *t* test analysis of global temperatures to evaluate a possible greenhouse effect (13.7, page 244); and a chi-square test of the survival rates of cabin and steerage passengers aboard the *Titanic* (19.14, page 384).
- Appendix B supplies answers to questions marked with asterisks. Other appendices provide a practical math review complete with self-diagnostic tests, a glossary of important terms, and tables for important statistical distributions.

INSTRUCTIONAL AIDS

An electronic version of an instructor's manual accompanies the text. The instructor's manual supplies answers omitted in the text (for about one-third of all questions), as well as sets of multiple-choice test items for each chapter, and a chapter-by-chapter commentary that reflects the authors' teaching experiences with this material. Instructors can access this material in the Instructor Companion Site at http://www.wiley.com/college/witte.

An electronic version of a student workbook, prepared by Beverly Dretzke of the University of Minnesota, also accompanies the text. Self-paced and self-correcting, the workbook contains problems, discussions, exercises, and tests that supplement the text. Students can access this material in the Student Companion Site at http://www.wiley.com/college/witte.

CHANGES IN THIS EDITION

- Update discussion of polling and random digit dialing in Section 8.4
- A new Section 14.11 on the "file drawer effect," whereby nonsignificant statistical findings are never published and the importance of replication.
- Updated numerical examples.
- New examples and questions throughout the book.
- Computer outputs and website have been updated.

USING THE BOOK

The book contains more material than is covered in most one-quarter or one-semester courses. Various chapters can be omitted without interrupting the main development. Typically, during a one-semester course we cover the entire book except for analysis of variance (Chapters 16, 17, and 18) and tests of ranked data (Chapter 20). An instructor who wishes to emphasize inferential statistics could skim some of the earlier chapters, particularly *Normal Distributions and Standard Scores (z)* (Chapter 5), and *Regression* (Chapter 7), while an instructor who desires a more applied emphasis could omit *Populations, Samples, and Probability* (Chapter 8) and *More about Hypothesis Testing* (Chapter 11).

ACKNOWLEDGMENTS

The authors wish to acknowledge their immediate family: Doris, Steve, Faith, Mike, Sharon, Andrea, Phil, Katie, Keegan, Camy, Brittany, Brent, Kristen, Scott, Joe, John, Jack, Carson, Sam, Margaret, Gretchen, Carrigan, Kedrick, and Alika. The first author also wishes to acknowledge his brothers and sisters: Henry, the late Lila, J. Stuart, A. Gerhart, and Etz; deceased parents: Henry and Emma; and all friends and relatives, past and present, including Arthur, Betty, Bob, Cal, David, Dick, Ellen, George, Grace, Harold, Helen, John, Joyce, Kayo, Kit, Mary, Paul, Ralph, Ruth, Shirley, and Suzanne.

Numerous helpful comments were made by those who reviewed the current and previous editions of this book: John W. Collins, Jr., Seton Hall University; Jelani Mandara, Northwestern University; L. E. Banderet, Northeastern University; S. Natasha Beretvas, University of Texas at Austin; Patricia M. Berretty, Fordham University; David Coursey, Florida State University; Shelia Kennison, Oklahoma State University; Melanie Kercher, Sam Houston State University; Jennifer H. Nolan, Loyola Marymount University; and Jonathan C. Pettibone, University of Alabama in Huntsville; Kevin Sumrall, Montgomery College; Sky Chafin, Grossmont College; Christine Ferri, Richard Stockton College of NJ; Ann Barich, Lewis University.

Special thanks to Carson Witte who proofread the entire manuscript twice.

Excellent editorial support was supplied by the people at John Wiley & Sons, Inc., most notably Abidha Sulaiman and Gladys Soto.

Contents

PREFACE iv

ACKNOWLEDGMENTS vi

1 INTRODUCTION 1

1.1 WHY STUDY STATISTICS? 2
1.2 WHAT IS STATISTICS? 2
1.3 MORE ABOUT INFERENTIAL STATISTICS 3
1.4 THREE TYPES OF DATA 6
1.5 LEVELS OF MEASUREMENT 7
1.6 TYPES OF VARIABLES 11
1.7 HOW TO USE THIS BOOK 15

Summary 16
Important Terms 17
Review Questions 17

PART 1 Descriptive Statistics: Organizing and Summarizing Data 21

2 DESCRIBING DATA WITH TABLES AND GRAPHS 22

TABLES (FREQUENCY DISTRIBUTIONS) 23

2.1 FREQUENCY DISTRIBUTIONS FOR QUANTITATIVE DATA 23
2.2 GUIDELINES 24
2.3 OUTLIERS 27
2.4 RELATIVE FREQUENCY DISTRIBUTIONS 28
2.5 CUMULATIVE FREQUENCY DISTRIBUTIONS 30
2.6 FREQUENCY DISTRIBUTIONS FOR QUALITATIVE (NOMINAL) DATA 31
2.7 INTERPRETING DISTRIBUTIONS CONSTRUCTED BY OTHERS 32

GRAPHS 33

2.8 GRAPHS FOR QUANTITATIVE DATA 33
2.9 TYPICAL SHAPES 37
2.10 A GRAPH FOR QUALITATIVE (NOMINAL) DATA 39
2.11 MISLEADING GRAPHS 40
2.12 DOING IT YOURSELF 41

Summary 42
Important Terms 43
Review Questions 43

3 DESCRIBING DATA WITH AVERAGES 47
3.1 MODE 48
3.2 MEDIAN 49
3.3 MEAN 51
3.4 WHICH AVERAGE? 53
3.5 AVERAGES FOR QUALITATIVE AND RANKED DATA 55
Summary 56
Important Terms 57
Key Equation 57
Review Questions 57

4 DESCRIBING VARIABILITY 60
4.1 INTUITIVE APPROACH 61
4.2 RANGE 62
4.3 VARIANCE 63
4.4 STANDARD DEVIATION 64
4.5 DETAILS: STANDARD DEVIATION 67
4.6 DEGREES OF FREEDOM (*df*) 75
4.7 INTERQUARTILE RANGE (IQR) 76
4.8 MEASURES OF VARIABILITY FOR QUALITATIVE AND RANKED DATA 78
Summary 78
Important Terms 79
Key Equations 79
Review Questions 79

5 NORMAL DISTRIBUTIONS AND STANDARD (*z*) SCORES 82
5.1 THE NORMAL CURVE 83
5.2 *z* SCORES 86
5.3 STANDARD NORMAL CURVE 87
5.4 SOLVING NORMAL CURVE PROBLEMS 89
5.5 FINDING PROPORTIONS 90
5.6 FINDING SCORES 95
5.7 MORE ABOUT *z* SCORES 100
Summary 103
Important Terms 103
Key Equations 103
Review Questions 103

6 DESCRIBING RELATIONSHIPS: CORRELATION 107
6.1 AN INTUITIVE APPROACH 108
6.2 SCATTERPLOTS 109
6.3 A CORRELATION COEFFICIENT FOR QUANTITATIVE DATA: *r* 113
6.4 DETAILS: COMPUTATION FORMULA FOR *r* 117
6.5 OUTLIERS AGAIN 118
6.6 OTHER TYPES OF CORRELATION COEFFICIENTS 119

6.7 COMPUTER OUTPUT 120
Summary 123
Important Terms and Symbols 124
Key Equations 124
Review Questions 124

7 REGRESSION 126
7.1 TWO ROUGH PREDICTIONS 127
7.2 A REGRESSION LINE 128
7.3 LEAST SQUARES REGRESSION LINE 130
7.4 STANDARD ERROR OF ESTIMATE, $s_{y|x}$ 133
7.5 ASSUMPTIONS 135
7.6 INTERPRETATION OF r^2 136
7.7 MULTIPLE REGRESSION EQUATIONS 141
7.8 REGRESSION TOWARD THE MEAN 141
Summary 143
Important Terms 144
Key Equations 144
Review Questions 144

PART 2 Inferential Statistics: Generalizing Beyond Data 147

8 POPULATIONS, SAMPLES, AND PROBABILITY 148

POPULATIONS AND SAMPLES 149
8.1 POPULATIONS 149
8.2 SAMPLES 150
8.3 RANDOM SAMPLING 151
8.4 TABLES OF RANDOM NUMBERS 151
8.5 RANDOM ASSIGNMENT OF SUBJECTS 153
8.6 SURVEYS OR EXPERIMENTS? 154

PROBABILITY 155
8.7 DEFINITION 155
8.8 ADDITION RULE 156
8.9 MULTIPLICATION RULE 157
8.10 PROBABILITY AND STATISTICS 161
Summary 162
Important Terms 163
Key Equations 163
Review Questions 163

9 SAMPLING DISTRIBUTION OF THE MEAN 168

9.1 WHAT IS A SAMPLING DISTRIBUTION? 169
9.2 CREATING A SAMPLING DISTRIBUTION FROM SCRATCH 170
9.3 SOME IMPORTANT SYMBOLS 173
9.4 MEAN OF ALL SAMPLE MEANS ($\mu_{\bar{X}}$) 173
9.5 STANDARD ERROR OF THE MEAN ($\sigma_{\bar{X}}$) 174
9.6 SHAPE OF THE SAMPLING DISTRIBUTION 176
9.7 OTHER SAMPLING DISTRIBUTIONS 178

Summary 178
Important Terms 179
Key Equations 179
Review Questions 179

10 INTRODUCTION TO HYPOTHESIS TESTING: THE *z* TEST 182

10.1 TESTING A HYPOTHESIS ABOUT SAT SCORES 183
10.2 *z* TEST FOR A POPULATION MEAN 185
10.3 STEP-BY-STEP PROCEDURE 186
10.4 STATEMENT OF THE RESEARCH PROBLEM 187
10.5 NULL HYPOTHESIS (H_0) 188
10.6 ALTERNATIVE HYPOTHESIS (H_1) 188
10.7 DECISION RULE 189
10.8 CALCULATIONS 190
10.9 DECISION 190
10.10 INTERPRETATION 191

Summary 191
Important Terms 192
Key Equations 192
Review Questions 193

11 MORE ABOUT HYPOTHESIS TESTING 195

11.1 WHY HYPOTHESIS TESTS? 196
11.2 STRONG OR WEAK DECISIONS 197
11.3 ONE-TAILED AND TWO-TAILED TESTS 199
11.4 CHOOSING A LEVEL OF SIGNIFICANCE (α) 202
11.5 TESTING A HYPOTHESIS ABOUT VITAMIN C 203
11.6 FOUR POSSIBLE OUTCOMES 204
11.7 IF H_0 REALLY IS TRUE 206
11.8 IF H_0 REALLY IS FALSE BECAUSE OF A *LARGE* EFFECT 207
11.9 IF H_0 REALLY IS FALSE BECAUSE OF A *SMALL* EFFECT 209
11.10 INFLUENCE OF SAMPLE SIZE 211
11.11 POWER AND SAMPLE SIZE 213

Summary 216
Important Terms 217
Review Questions 218

12 ESTIMATION (CONFIDENCE INTERVALS) 221

12.1 POINT ESTIMATE FOR μ 222
12.2 CONFIDENCE INTERVAL (CI) FOR μ 222
12.3 INTERPRETATION OF A CONFIDENCE INTERVAL 226
12.4 LEVEL OF CONFIDENCE 226
12.5 EFFECT OF SAMPLE SIZE 227
12.6 HYPOTHESIS TESTS OR CONFIDENCE INTERVALS? 228
12.7 CONFIDENCE INTERVAL FOR POPULATION PERCENT 228
Summary 230
Important Terms 230
Key Equation 230
Review Questions 231

13 *t* TEST FOR ONE SAMPLE 233

13.1 GAS MILEAGE INVESTIGATION 234
13.2 SAMPLING DISTRIBUTION OF t 234
13.3 t TEST 237
13.4 COMMON THEME OF HYPOTHESIS TESTS 238
13.5 REMINDER ABOUT DEGREES OF FREEDOM 238
13.6 DETAILS: ESTIMATING THE STANDARD ERROR ($s_{\overline{X}}$) 238
13.7 DETAILS: CALCULATIONS FOR THE t TEST 239
13.8 CONFIDENCE INTERVALS FOR μ BASED ON t 241
13.9 ASSUMPTIONS 242
Summary 242
Important Terms 243
Key Equations 243
Review Questions 243

14 *t* TEST FOR TWO INDEPENDENT SAMPLES 245

14.1 EPO EXPERIMENT 246
14.2 STATISTICAL HYPOTHESES 247
14.3 SAMPLING DISTRIBUTION OF $\overline{X}_1 - \overline{X}_2$ 248
14.4 t TEST 250
14.5 DETAILS: CALCULATIONS FOR THE t TEST 252
14.6 p-VALUES 255
14.7 STATISTICALLY SIGNIFICANT RESULTS 258
14.8 ESTIMATING EFFECT SIZE: POINT ESTIMATES AND CONFIDENCE INTERVALS 259
14.9 ESTIMATING EFFECT SIZE: COHEN'S d 262
14.10 META-ANALYSIS 264
14.11 IMPORTANCE OF REPLICATION 264
14.12 REPORTS IN THE LITERATURE 265

14.13 ASSUMPTIONS 266
14.14 COMPUTER OUTPUT 267
Summary 268
Important Terms 268
Key Equations 269
Review Questions 269

15 *t* TEST FOR TWO RELATED SAMPLES (REPEATED MEASURES) 273
15.1 EPO EXPERIMENT WITH REPEATED MEASURES 274
15.2 STATISTICAL HYPOTHESES 277
15.3 SAMPLING DISTRIBUTION OF $\bar{D}$ 277
15.4 *t* TEST 278
15.5 DETAILS: CALCULATIONS FOR THE *t* TEST 279
15.6 ESTIMATING EFFECT SIZE 281
15.7 ASSUMPTIONS 283
15.8 OVERVIEW: THREE *t* TESTS FOR POPULATION MEANS 283
15.9 *t* TEST FOR THE POPULATION CORRELATION COEFFICIENT, ρ 285
Summary 287
Important Terms 288
Key Equations 288
Review Questions 288

16 ANALYSIS OF VARIANCE (ONE FACTOR) 292
16.1 TESTING A HYPOTHESIS ABOUT SLEEP DEPRIVATION AND AGGRESSION 293
16.2 TWO SOURCES OF VARIABILITY 294
16.3 *F* TEST 296
16.4 DETAILS: VARIANCE ESTIMATES 299
16.5 DETAILS: MEAN SQUARES (*MS*) AND THE *F* RATIO 304
16.6 TABLE FOR THE *F* DISTRIBUTION 305
16.7 ANOVA SUMMARY TABLES 307
16.8 *F* TEST IS NONDIRECTIONAL 308
16.9 ESTIMATING EFFECT SIZE 308
16.10 MULTIPLE COMPARISONS 311
16.11 OVERVIEW: FLOW CHART FOR ANOVA 315
16.12 REPORTS IN THE LITERATURE 315
16.13 ASSUMPTIONS 316
16.14 COMPUTER OUTPUT 316
Summary 317
Important Terms 318
Key Equations 318
Review Questions 319

17 ANALYSIS OF VARIANCE (REPEATED MEASURES) 322
17.1 SLEEP DEPRIVATION EXPERIMENT WITH REPEATED MEASURES 323
17.2 *F* TEST 324

17.3 TWO COMPLICATIONS 325
17.4 DETAILS: VARIANCE ESTIMATES 326
17.5 DETAILS: MEAN SQUARE (*MS*) AND THE *F* RATIO 329
17.6 TABLE FOR *F* DISTRIBUTION 331
17.7 ANOVA SUMMARY TABLES 331
17.8 ESTIMATING EFFECT SIZE 333
17.9 MULTIPLE COMPARISONS 333
17.10 REPORTS IN THE LITERATURE 335
17.11 ASSUMPTIONS 336
Summary 336
Important Terms 336
Key Equations 337
Review Questions 337

18 ANALYSIS OF VARIANCE (TWO FACTORS) 339
18.1 A TWO-FACTOR EXPERIMENT: RESPONSIBILITY IN CROWDS 340
18.2 THREE *F* TESTS 342
18.3 INTERACTION 344
18.4 DETAILS: VARIANCE ESTIMATES 347
18.5 DETAILS: MEAN SQUARES (*MS*) AND *F* RATIOS 351
18.6 TABLE FOR THE *F* DISTRIBUTION 353
18.7 ESTIMATING EFFECT SIZE 353
18.8 MULTIPLE COMPARISONS 354
18.9 SIMPLE EFFECTS 355
18.10 OVERVIEW: FLOW CHART FOR TWO-FACTOR ANOVA 358
18.11 REPORTS IN THE LITERATURE 358
18.12 ASSUMPTIONS 360
18.13 OTHER TYPES OF ANOVA 360
Summary 360
Important Terms 361
Key Equations 361
Review Questions 361

19 CHI-SQUARE (χ^2) TEST FOR QUALITATIVE (NOMINAL) DATA 365

ONE-VARIABLE χ^2 TEST 366
19.1 SURVEY OF BLOOD TYPES 366
19.2 STATISTICAL HYPOTHESES 366
19.3 DETAILS: CALCULATING χ^2 367
19.4 TABLE FOR THE χ^2 DISTRIBUTION 369
19.5 χ^2 TEST 370

TWO-VARIABLE χ^2 TEST 372
19.6 LOST LETTER STUDY 372
19.7 STATISTICAL HYPOTHESES 373
19.8 DETAILS: CALCULATING χ^2 373

19.9 TABLE FOR THE χ^2 DISTRIBUTION 376
19.10 χ^2 TEST 376
19.11 ESTIMATING EFFECT SIZE 377
19.12 ODDS RATIOS 378
19.13 REPORTS IN THE LITERATURE 380
19.14 SOME PRECAUTIONS 380
19.15 COMPUTER OUTPUT 381
Summary 382
Important Terms 382
Key Equations 382
Review Questions 382

20 TESTS FOR RANKED (ORDINAL) DATA 386
20.1 USE ONLY WHEN APPROPRIATE 387
20.2 A NOTE ON TERMINOLOGY 387
20.3 MANN–WHITNEY *U* TEST (TWO INDEPENDENT SAMPLES) 387
20.4 WILCOXON *T* TEST (TWO RELATED SAMPLES) 392
20.5 KRUSKAL–WALLIS *H* TEST (THREE OR MORE INDEPENDENT SAMPLES) 396
20.6 GENERAL COMMENT: TIES 400
Summary 400
Important Terms 400
Review Questions 400

21 POSTSCRIPT: WHICH TEST? 403
21.1 DESCRIPTIVE OR INFERENTIAL STATISTICS? 404
21.2 HYPOTHESIS TESTS OR CONFIDENCE INTERVALS? 404
21.3 QUANTITATIVE OR QUALITATIVE DATA? 404
21.4 DISTINGUISHING BETWEEN THE TWO TYPES OF DATA 406
21.5 ONE, TWO, OR MORE GROUPS? 407
21.6 CONCLUDING COMMENTS 408
Review Questions 408

APPENDICES 411
A MATH REVIEW 411
B ANSWERS TO SELECTED QUESTIONS 419
C TABLES 457
D GLOSSARY 471

INDEX 477

STATISTICS

Eleventh Edition

CHAPTER 1 Introduction

1.1 WHY STUDY STATISTICS?
1.2 WHAT IS STATISTICS?
1.3 MORE ABOUT INFERENTIAL STATISTICS
1.4 THREE TYPES OF DATA
1.5 LEVELS OF MEASUREMENT
1.6 TYPES OF VARIABLES
1.7 HOW TO USE THIS BOOK

Summary / Important Terms / Review Questions

Preview

Statistics deals with variability. You're different from everybody else (and, we hope, proud of it). Today differs from both yesterday and tomorrow. In an experiment designed to detect whether psychotherapy improves self-esteem, self-esteem scores will differ among subjects in the experiment, whether or not psychotherapy improves self-esteem.

Beginning with Chapter 2, ***descriptive statistics*** *will provide tools, such as tables, graphs, and averages, that help you describe and organize the inevitable variability among observations. For example, self-esteem scores (on a scale of 0 to 50) for a group of college students might approximate a bell-shaped curve with an average score of 32 and a range of scores from 18 to 49.*

Beginning with Chapter 8, ***inferential statistics*** *will supply powerful concepts that, by adjusting for the pervasive effects of variability, permit you to generalize beyond limited sets of observations. For example, inferential statistics might help us decide whether—after an adjustment has been made for background variability (or chance)—an observed improvement in self-esteem scores can be attributed to psychotherapy rather than to chance.*

Chapter 1 provides an overview of both descriptive and inferential statistics, and it also introduces a number of terms—some from statistics and some from math and research methods—with which you already may have some familiarity. These terms will clarify a number of important distinctions that will aid your progress through the book.

1.1 WHY STUDY STATISTICS?

You're probably taking a statistics course because it's required, and your feelings about it may be more negative than positive. Let's explore some of the reasons why you should study statistics. For instance, recent issues of a daily newspaper carried these items:

- The annual earnings of college graduates exceed, *on average,* those of high school graduates by $20,000.
- On the basis of existing research, there is *no evidence of a relationship* between family size and the scores of adolescents on a test of psychological adjustment.
- Heavy users of tobacco suffer *significantly* more respiratory ailments than do nonusers.

Having learned some statistics, you'll not stumble over the italicized phrases. Nor, as you continue reading, will you hesitate to probe for clarification by asking, "Which *average* shows higher annual earnings?" or "What constitutes a *lack of evidence* about a relationship?" or "How many more is *significantly more* respiratory ailments?"

A statistical background is indispensable in understanding research reports within your special area of interest. Statistical references punctuate the results sections of most research reports. Often expressed with parenthetical brevity, these references provide statistical support for the researcher's conclusions:

- Subjects who engage in daily exercise score higher on tests of self-esteem than do subjects who do not exercise [$p < .05$].
- Highly anxious students are perceived by others as less attractive than nonanxious students [$t\ (48) = 3.21$, $p < .01$, $d = .42$].
- Attitudes toward extramarital sex depend on socioeconomic status [$x^2\ (4, n = 185) = 11.49$, $p < .05$, $\phi^2_c = .03$].

Having learned some statistics, you will be able to decipher the meaning of these symbols and consequently read these reports more intelligently.

Sometime in the future—possibly sooner than you think—you might want to plan a statistical analysis for a research project of your own. Having learned some statistics, you'll be able to plan the statistical analysis for modest projects involving straightforward research questions. If your project requires more advanced statistical analysis, you'll know enough to consult someone with more training in statistics. Once you begin to understand basic statistical concepts, you will discover that, with some guidance, your own efforts often will enable you to use and interpret more advanced statistical analysis required by your research.

1.2 WHAT IS STATISTICS?

It is difficult to imagine, even as a fantasy exercise, a world where there is no variability—where, for example, everyone has the same physical characteristics, intelligence, attitudes, etc. Knowing that one person is 70 inches tall, and has an intelligence quotient (IQ) of 125 and a favorable attitude toward capital punishment, we could immediately conclude that everyone else also has these characteristics. This mind-numbing world would have little to recommend it, other than that there would be no need for the field of statistics (and a few of us probably would be looking for work).

Descriptive Statistics

Statistics exists because of the prevalence of variability in the real world. In its simplest form, known as **descriptive statistics,** *statistics provides us with tools—tables, graphs, averages, ranges, correlations—for organizing and summarizing the inevitable variability in collections of actual observations or scores.* Examples are:

1. A tabular listing, ranked from most to least, of the total number of romantic affairs during college reported anonymously by each member of your stat class
2. A graph showing the annual change in global temperature during the last 30 years
3. A report that describes the average difference in grade point average (GPA) between college students who regularly drink alcoholic beverages and those who don't

Inferential Statistics

Statistics also provides tools—a variety of tests and estimates—for generalizing beyond collections of actual observations. This more advanced area is known as **inferential statistics.** Tools from inferential statistics permit us to use a relatively small collection of actual observations to evaluate, for example:

1. A pollster's claim that a majority of all U.S. voters favor stronger gun control laws
2. A researcher's hypothesis that, on average, meditators report fewer headaches than do nonmeditators
3. An assertion about the relationship between job satisfaction and overall happiness

In this book, you will encounter the most essential tools of descriptive statistics (Part 1), beginning with Chapter 2, and those of inferential statistics (Part 2), beginning with Chapter 8.

Progress Check *1.1 Indicate whether each of the following statements typifies descriptive statistics (because it describes sets of actual observations) or inferential statistics (because it generalizes beyond sets of actual observations).

(a) Students in my statistics class are, on average, 23 years old.

(b) The population of the world exceeds 7 billion (that is, 7,000,000,000 or 1 million multiplied by 7000).

(c) Either four or eight years have been the most frequent terms of office actually served by U.S. presidents.

(d) Sixty-four percent of all college students favor right-to-abortion laws.

Answers on page 420.

1.3 MORE ABOUT INFERENTIAL STATISTICS

Population

Any complete collection of observations or potential observations.

Populations and Samples

Inferential statistics is concerned with generalizing beyond sets of actual observations, that is, with generalizing from a sample to a population. In statistics, a **population**

Sample
Any smaller collection of actual observations from a population.

refers to *any complete collection of observations or potential observations,* whereas a **sample** refers to *any smaller collection of actual observations drawn from a population.* In everyday life, populations often are viewed as collections of real objects (e.g., people, whales, automobiles), whereas in statistics, populations may be viewed more abstractly as collections of properties or measurements (e.g., the ethnic backgrounds of people, life spans of whales, gas mileage of automobiles).

Depending on your perspective, a given set of observations can be either a population or a sample. For instance, the weights reported by 53 male statistics students in **Table 1.1** can be viewed either as a population, because you are concerned about exceeding the load-bearing capacity of an excursion boat (chartered by the 53 students to celebrate successfully completing their stat class!), or as a sample from a population because you wish to generalize to the weights of *all* male statistics students or *all* male college students.

Table 1.1
QUANTITATIVE DATA: WEIGHTS (IN POUNDS) OF MALE STATISTICS STUDENTS

160	168	133	170	150	165	158	165
193	169	245	160	152	190	179	157
226	160	170	180	150	156	190	156
157	163	152	158	225	135	165	135
180	172	160	170	145	185	152	
205	151	220	166	152	159	156	
165	157	190	206	172	175	154	

Ordinarily, populations are quite large and exist only as potential observations (e.g., the *potential* scores of all U.S. college students on a test that measures anxiety). On the other hand, samples are relatively small and exist as actual observations (the *actual* scores of 100 college students on the test for anxiety). When using a sample (100 actual scores) to generalize to a population (millions of potential scores), it is important that the sample represent the population; otherwise, any generalization might be erroneous. Although conveniently accessible, the anxiety test scores for the 100 students in stat classes at your college probably would not be representative of the scores for all students. If you think about it, these 100 stat students might tend to have either higher or lower anxiety scores than those in the target population for numerous reasons including, for instance, the fact that the 100 students are mostly psychology majors enrolled in a required stat class at your particular college.

Random Sampling (Surveys)

Random Sampling
A procedure designed to ensure that each potential observation in the population has an equal chance of being selected in a survey.

Whenever possible, a sample should be randomly selected from a population in order to increase the likelihood that the sample accurately represents the population. **Random sampling** *is a procedure designed to ensure that each potential observation in the population has an equal chance of being selected in a survey.* Classic examples of random samples are a state lottery where each number from 1 to 99 in the population has an equal chance of being selected as one of the five winning numbers or a nationwide opinion survey in which each telephone number has an equal chance of being selected as a result of a series of random selections, beginning with a three-digit area code and ending with a specific seven-digit telephone number.

Random sampling can be very difficult when a population lacks structure (e.g., all persons currently in psychotherapy) or specific boundaries (e.g., all volunteers who could conceivably participate in an experiment). In this case, a random sample

becomes an ideal that can only be approximated—always with an effort to remove obvious biases that might cause the sample to misrepresent the population. For example, lacking the resources to sample randomly the target population of all U.S. college students, you might obtain scores by randomly selecting the 100 students, not just from stat classes at your college but also from one or more college directories, possibly using some of the more elaborate techniques described in Chapter 8. Insofar as your sample only approximates a true random sample, any resulting generalizations should be qualified. For example, if the 100 students were randomly selected only from several public colleges in northern California, this fact should be noted, and any generalizations to all college students in the United States would be both provisional and open to criticism.

Random Assignment (Experiments)

Estimating the average anxiety score for all college students probably would not generate much interest. Instead, we might be interested in determining whether relaxation training causes, on average, a reduction in anxiety scores between two groups of otherwise similar college students. Even if relaxation training has no effect on anxiety scores, we would expect average scores for the two groups to differ because of the inevitable variability between groups. The question becomes: How should we interpret the apparent difference between the treatment group and the control group? Once variability has been taken into account, should the difference be viewed as real (and attributable to relaxation training) or as transitory (and merely attributable to variability or chance)?

College students in the relaxation experiment probably are not a random sample from any intact population of interest, but rather a *convenience sample* consisting of volunteers from a limited pool of students fulfilling a course requirement. Accordingly, our focus shifts from random sampling to the random assignment of volunteers to the two groups. **Random assignment** *signifies that each person has an equal chance of being assigned to any group in an experiment.* Using procedures described in Chapter 8, random assignment should be employed whenever possible. Because chance dictates the membership of both groups, not only does random assignment minimize any biases that might favor one group or another, it also serves as a basis for estimating the role of variability in any observed result. Random assignment allows us to evaluate any finding, such as the actual average difference between two groups, to determine whether this difference is larger than expected just by chance, once variability is taken into account. In other words, it permits us to generalize beyond mere appearances and determine whether the average difference merits further attention because it *probably is real* or whether it should be ignored because it *can be attributed to variability or chance.*

Random Assignment
A procedure designed to ensure that each person has an equal chance of being assigned to any group in an experiment.

Overview: Surveys and Experiments

Figure 1.1 compares surveys and experiments. Based on random samples from populations, surveys permit generalizations from samples back to populations. Based on the random assignment of volunteers to groups, experiments permit decisions about whether differences between groups are real or merely transitory.

PROGRESS CHECK *1.2 Indicate whether each of the following terms is associated primarily with a survey (S) or an experiment (E).

(a) random assignment

(b) representative

(c) generalization to the population

(d) control group

Answers on page 420.

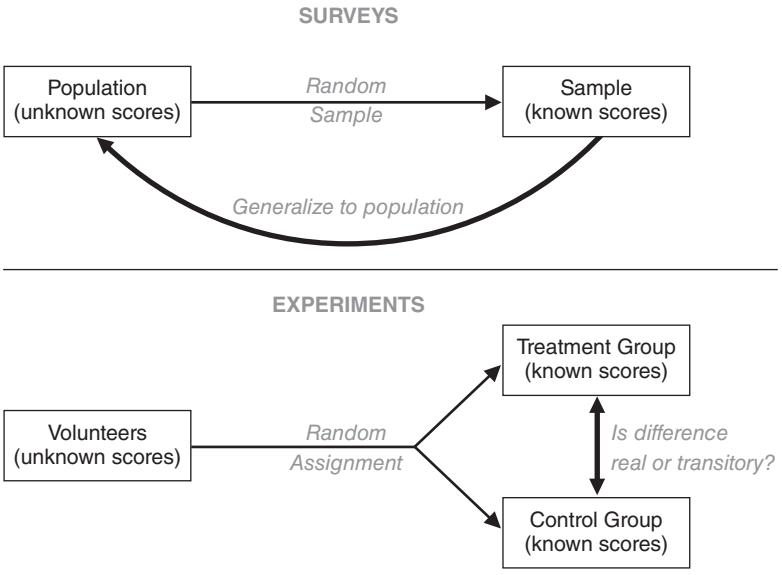


FIGURE 1.1
Overview: surveys and experiments.

(e) real difference

(f) random selection

(g) convenience sample

(h) volunteers

Answers on page 420.

Data
A collection of actual observations or scores in a survey or an experiment

1.4 THREE TYPES OF DATA

Any statistical analysis is performed on **data,** a *collection of actual observations or scores in a survey or an experiment.*

The precise form of a statistical analysis often depends on whether data are qualitative, ranked, or quantitative.

Qualitative Data
A set of observations where any single observation is a word, letter, or numerical code that represents a class or category.

Ranked Data
A set of observations where any single observation is a number that indicates relative standing.

Quantitative Data
A set of observations where any single observation is a number that represents an amount or a count.

Generally, **qualitative data** consist of words (Yes or No), letters (Y or N), or numerical codes (0 or 1) that represent a class or category. **Ranked data** consist of numbers (1st, 2nd, . . . 40th place) that represent relative standing within a group. **Quantitative data** consist of numbers (weights of 238, 170, . . . 185 lbs) that represent an amount or a count. To determine the type of data, focus on a single observation in any collection of observations. For example, the weights reported by 53 male students in Table 1.1 are quantitative data, since any single observation, such as 160 lbs, represents an amount of weight. If the weights in Table 1.1 had been replaced with ranks, beginning with a rank of 1 for the lightest weight of 133 lbs and ending with a rank of 53 for the heaviest weight of 245 lbs, these numbers would have been ranked data, since any single observation represents not an amount, but only relative standing within the group of 53 students. Finally, the Y and N replies of students in **Table 1.2** are qualitative data, since any single observation is a letter that represents a class of replies.

Table 1.2
QUALITATIVE DATA: "DO YOU HAVE A FACEBOOK PROFILE?" YES (Y) OR NO (N) REPLIES OF STATISTICS STUDENTS

Y	Y	Y	N	N	Y	Y	Y
Y	Y	Y	N	N	Y	Y	Y
N	Y	N	Y	Y	Y	Y	Y
Y	Y	N	Y	N	Y	N	Y
Y	N	Y	N	N	Y	Y	Y
Y	Y	N	Y	Y	Y	Y	Y
N	N	N	N	Y	N	N	Y
Y	Y	Y	Y	Y	N	Y	N
Y	Y	Y	Y	N	N	Y	Y
N	Y	N	N	Y	Y	Y	Y
	Y	Y	N				

Progress Check *1.3 Indicate whether each of the following terms is *qualitative* (because it's a word, letter, or numerical code representing a class or category); *ranked* (because it's a number representing relative standing); or *quantitative* (because it's a number representing an amount or a count).

(a) ethnic group

(b) age

(c) family size

(d) academic major

(e) sexual preference

(f) IQ score

(g) net worth (dollars)

(h) third-place finish

(i) gender

(j) temperature

Answers on page 420.

Level of Measurement

Specifies the extent to which a number (or word or letter) actually represents some attribute and, therefore, has implications for the appropriateness of various arithmetic operations and statistical procedures.

1.5 LEVELS OF MEASUREMENT

Learned years ago in grade school, the abstract statement that $2 + 2 = 4$ qualifies as one of life's everyday certainties, along with taxes and death. However, not all numbers have the same interpretation. For instance, it wouldn't make sense to find the sum of two Social Security numbers or to claim that, when viewed as indicators of academic achievement, two GPAs of 2.0 equal a GPA of 4.0. To clarify further the differences among the three types of data, let's introduce the notion of level of measurement. Looming behind any data, the **level of measurement** *specifies the extent to which a number (or word or letter) actually represents some attribute and, therefore, has implications for the appropriateness of various arithmetic operations and statistical procedures.*

For our purposes, there are three levels of measurement—nominal, ordinal, and interval/ratio—and these levels are paired with qualitative, ranked, and quantitative data, respectively. The properties of these levels—and the usefulness of their associated numbers—vary from nominal, the simplest level with only one property, to interval/ratio, the most complex level with four properties. Progressively more complex levels contain all properties of simpler levels, plus one or two new properties.

More complex levels of measurement are associated with numbers that, because they better represent attributes, permit a wider variety of arithmetic operations and statistical procedures.

Qualitative Data and Nominal Measurement

Nominal Measurement
Words, letters, or numerical codes of qualitative data that reflect differences in kind based on classification.

If people are classified as either male or female (or coded as 1 or 2), the data are qualitative and measurement is nominal. *The single property of* **nominal measurement** *is classification*—that is, sorting observations into different classes or categories. Words, letters, or numerical codes reflect only differences in kind, not differences in amount. Examples of nominal measurement include classifying mood disorders as manic, bipolar, or depressive; sexual preferences as heterosexual, homosexual, bisexual, or nonsexual; and attitudes toward stricter pollution controls as favor, oppose, or undecided.

A distinctive feature of nominal measurement is its bare-bones representation of any attribute. For instance, a student is either male or female. Even with the introduction of arbitrary numerical codes, such as 1 for male and 2 for female, it would never be appropriate to claim that, because female is 2 and male is 1, females have twice as much gender as males. Similarly, calculating an average with these numbers would be meaningless. Because of these limitations, only a few sections of this book and Chapter 19 are dedicated exclusively to an analysis of qualitative data with nominal measurement.

Ranked Data and Ordinal Measurement

Ordinal Measurement
Relative standing of ranked data that reflects differences in degree based on order.

When any single number indicates only *relative standing,* such as first, second, or tenth place in a horse race or in a class of graduating seniors, the data are *ranked* and the level of measurement is ordinal. *The distinctive property of* **ordinal measurement** *is order*. Comparatively speaking, a first-place finish reflects the fastest finish in a horse race or the highest GPA among graduating seniors. Although first place in a horse race indicates a *faster* finish than second place, we don't know *how much* faster.

Since ordinal measurement fails to reflect the actual distance between adjacent ranks, simple arithmetic operations with ranks are inappropriate. For example, it's inappropriate to conclude that the arithmetic mean of ranks 1 and 3 equals rank 2, since this assumes that the actual distance between ranks 1 and 2 equals the distance between ranks 2 and 3. Instead, these distances might be very different. For example, rank 2 might be virtually tied with either rank 1 or rank 3. Only a few sections of this book and Chapter 20 are dedicated exclusively to an analysis of ranked data with ordinal measurement.*

*Strictly speaking, ordinal measurement also can be associated with qualitative data whose classes are ordered. Examples of *ordered* qualitative data include the classification of skilled workers as master craftsman, journeyman, or apprentice; socioeconomic status as low, middle, or high; and academic grades as A, B, C, D, or F. It's worth distinguishing between qualitative data with nominal and ordinal measurement because, as described in Chapters 3 and 4, a few extra statistical procedures are available for ordered qualitative data.

Quantitative Data and Interval/Ratio Measurement

Interval/Ratio Measurement
Amounts or counts of quantitative data reflect differences in degree based on equal intervals and a true zero.

Often the products of familiar measuring devices, such as rulers, clocks, or meters, *the distinctive properties of* **interval/ratio measurement** *are equal intervals and a true zero.* Weighing yourself on a bathroom scale qualifies as interval/ratio measurement. *Equal intervals* imply that hefting a 10-lb weight while on the bathroom scale always registers your actual weight plus 10 lbs. Equal intervals imply that the difference between 120 and 130 lbs represents an *amount* of weight equal to the difference between 130 and 140 lbs, and it's appropriate to describe one person's weight as a certain amount greater than another's.

A *true zero* signifies that the bathroom scale registers 0 when not in use—that is, when weight is completely absent. Since the bathroom scale possesses a true zero, numerical readings reflect the *total amount* of a person's weight, and it's appropriate to describe one person's weight as a certain ratio of another's. It can be said that the weight of a 140-lb person is twice that of a 70-lb person.

In the absence of a true zero, numbers—much like the exposed tips of icebergs—fail to reflect the total amount being measured. For example, a reading of 0 on the Fahrenheit temperature scale does not reflect the complete absence of heat—that is, the absence of any molecular motion. In fact, true zero equals –459.4°F on this scale. It would be inappropriate, therefore, to claim that 80°F is twice as hot as 40°F. An appropriate claim could be salvaged by adding 459.4°F to each of these numbers: 80° becomes 539.4° and 40° becomes 499.4°. Clearly, 539.4°F is not twice as hot as 499.4°F.

Interval/ratio measurement appears in the behavioral and social sciences as, for example, bar-press rates of rats in Skinner boxes; the minutes of dream-friendly rapid eye movement (REM) sleep among participants in a sleep-deprivation experiment; and the total number of eye contacts during verbal disputes between romantically involved couples. Thanks to the considerable amount of information conveyed by each observation, interval/ratio measurement permits meaningful arithmetic operations, such as calculating arithmetic means, as well as the many statistical procedures for quantitative data described in this book.

Measurement of Nonphysical Characteristics

When numbers represent nonphysical characteristics, such as intellectual aptitude, psychopathic tendency, or emotional maturity, the attainment of interval/ratio measurement often is questionable. For example, there is no external standard (such as the 10-lb weight) to demonstrate that the addition of a fixed amount of intellectual aptitude always produces an equal increase in IQ scores (equal intervals). There also is no instrument (such as the unoccupied bathroom scale) that registers an IQ score of 0 when intellectual aptitude is completely absent (true zero).

In the absence of equal intervals, it would be inappropriate to claim that the difference between IQ scores of 120 and 130 represents the same amount of intellectual aptitude as the difference between IQ scores of 130 and 140. Likewise, in the absence of a true zero, it would be inappropriate to claim that an IQ score of 140 represents twice as much intellectual aptitude as an IQ score of 70.

Other interpretations are possible. One possibility is to treat IQ scores as attaining only ordinal measurement—that is, for example, a score of 140 represents more intellectual aptitude than a score of 130—without specifying the actual size of this difference. This strict interpretation would greatly restrict the number of statistical procedures for use with behavioral and social data. A looser (and much more common) interpretation, adopted in this book, assumes that, although lacking a true zero, IQ scores provide a crude measure of corresponding differences in intellectual aptitude (equal intervals). Thus, the difference between IQ scores of 120 and 130 represents a *roughly similar* amount of intellectual aptitude as the difference between scores of 130 and 140.