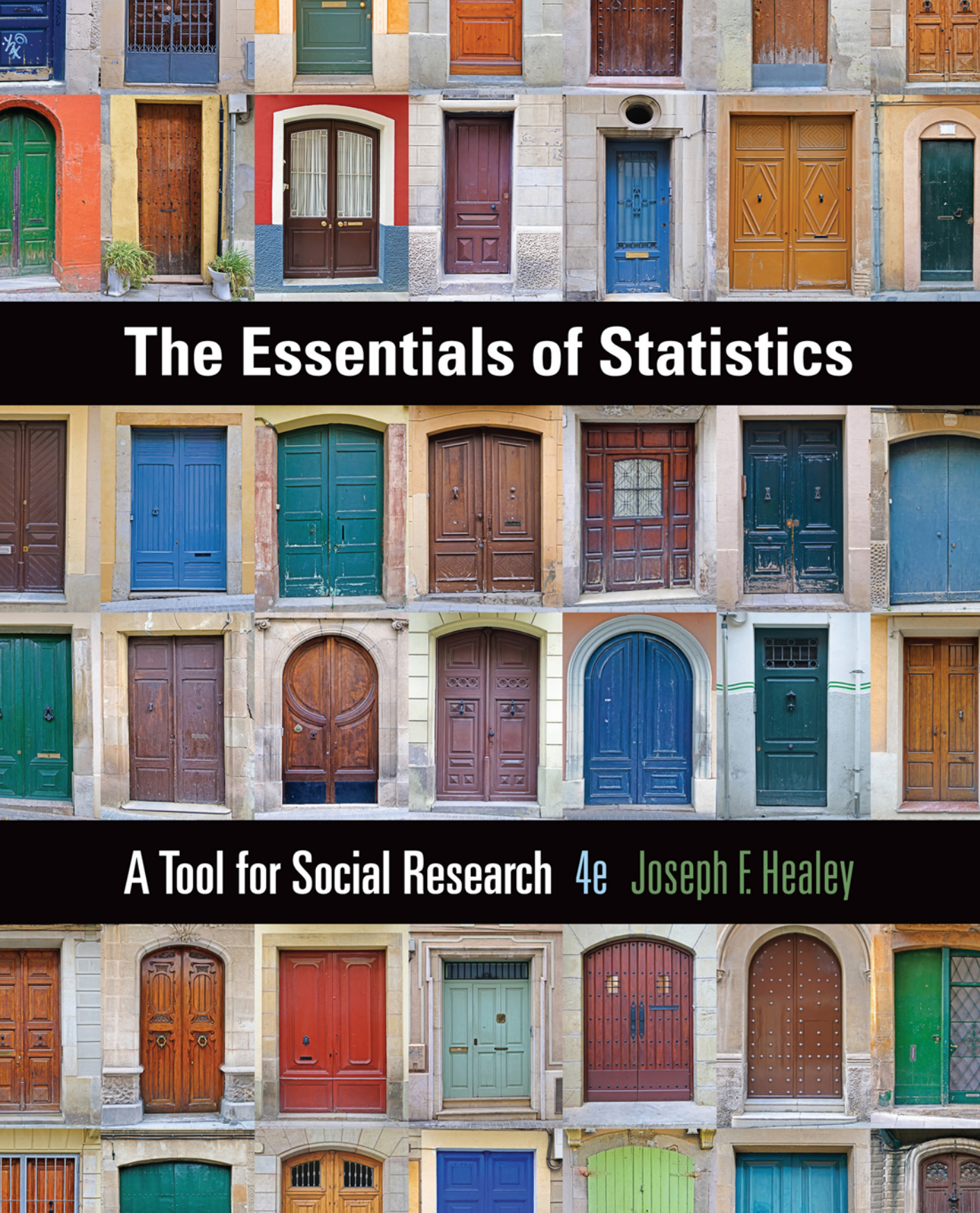

The Essentials of Statistics

A Tool for Social Research 4e

Joseph F. Healey

THE ESSENTIALS OF STATISTICS

A Tool for Social Research

Fourth Edition

Joseph F. Healey

Christopher Newport University

CENGAGE Learning®

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

CENGAGE Learning®

***The Essentials of Statistics: A Tool for Social Research*, Fourth Edition**
Joseph F. Healey

Product Director: Marta Lee-Perriard
Product Manager: Jennifer Harrison
Content Developer: Lori Bradshaw
Product Assistant: Chelsea Meredith
Media Developer: John Chell
Marketing Manager: Kara Kindstrom
Content Project Manager: Cheri Palmer
Art Director: Michelle Kunkler
Manufacturing Planner: Judy Inouye
Production, Composition, and Illustration: MPS Limited
Text Researcher: Lumina Datamatics
Copy Editor: MPS Limited
Text Designer: Diane Beasley
Cover Designer: Ramsdell Design, LLC
Cover Image: © matthi/Shutterstock.com

Library of Congress Control Number: 2014959328

ISBN: 978-1-305-09383-6

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at **www.cengage.com/global**.

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com**.

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**.

Printed in the United States of America
Print Number: 01 Print Year: 2015

Brief Contents

Preface xiii
Prologue: Basic Mathematics Review xx
Chapter 1 Introduction 1

Part I Descriptive Statistics 23

Chapter 2 Basic Descriptive Statistics: Tables, Percentages, Ratios and Rates, and Graphs 24
Chapter 3 Measures of Central Tendency 63
Chapter 4 Measures of Dispersion 87
Chapter 5 The Normal Curve 118

Part II Inferential Statistics 139

Chapter 6 Introduction to Inferential Statistics, the Sampling Distribution, and Estimation 140
Chapter 7 Hypothesis Testing I: The One-Sample Case 176
Chapter 8 Hypothesis Testing II: The Two-Sample Case 207
Chapter 9 Hypothesis Testing III: The Analysis of Variance 238
Chapter 10 Hypothesis Testing IV: Chi Square 265

Part III Bivariate Measures of Association 291

Chapter 11 Bivariate Association for Nominal- and Ordinal-Level Variables 292
Chapter 12 Association Between Variables Measured at the Interval-Ratio Level 328

Part IV Multivariate Techniques 363

Chapter 13 Partial Correlation and Multiple Regression and Correlation 364

Appendix A Area Under the Normal Curve 395
Appendix B Distribution of *t* 399
Appendix C Distribution of Chi Square 400
Appendix D Distribution of *F* 401
Appendix E Using Statistics: Ideas for Research Projects 403
Appendix F An Introduction to SPSS 409

Appendix G Codebooks for All Data Sets 417
Answers to Odd-Numbered End-of-Chapter Problems 429
Glossary 441
Index 447

Detailed Contents

Preface xiii

Prologue: Basic Mathematics Review xx

Chapter 1 Introduction 1

Why Study Statistics? 2
The Role of Statistics in Scientific Inquiry 2
The Goals of This Text 6
Descriptive and Inferential Statistics 7
Level of Measurement 10
Becoming a Critical Consumer: Introduction 14
One Step at a Time: Determining the Level of Measurement of a Variable 14
Some Final Points 15
SUMMARY / 17 • GLOSSARY / 17 • PROBLEMS / 18
YOU ARE THE RESEARCHER: INTRODUCTION 21

Part I Descriptive Statistics 23

Chapter 2 Basic Descriptive Statistics: Tables, Percentages, Ratios and Rates, and Graphs 24

Frequency Distributions for Nominal-Level Variables 25
Enhancing Clarity: Percentages and Proportions 26
One Step at a Time: Finding Percentages and Proportions 28
Applying Statistics 2.1: Using Percentages and Proportions 29
Applying Statistics 2.2: Communicating with Statistics 29
Frequency Distributions for Ordinal-Level Variables 30
Frequency Distributions for Interval-Ratio-Level Variables 31
One Step at a Time: Finding Midpoints 34
One Step at a Time: Adding Cumulative Frequency and Percentage Columns to Frequency Distributions 35
One Step at a Time: Constructing Frequency Distributions for Interval-Ratio Variables 37
Applying Statistics 2.3: Frequency Distributions 38
Using SPSS to Produce Frequency Distributions 38
Ratios, Rates, and Percentage Change 40

Applying Statistics 2.4: Ratio 41
Applying Statistics 2.5: Rates 42
Applying Statistics 2.6: Percentage Change 44
One Step at a Time: Finding Ratios, Rates, and Percent Change 45
Becoming a Critical Consumer: Lying with Statistics 46
Using Graphs to Present Data 46
Becoming a Critical Consumer: Tricks with Graphs 50
Using SPSS to Produce Graphs 51
Becoming a Critical Consumer: Statistics in the Professional Literature 53
SUMMARY / 54 • SUMMARY OF FORMULAS / 55 •
GLOSSARY / 55 • PROBLEMS / 55
YOU ARE THE RESEARCHER: IS THERE A "CULTURE WAR"
IN THE UNITED STATES? 60

Chapter 3 Measures of Central Tendency 63

The Mode 64
The Median 66
One Step at a Time: Finding the Median 67
The Mean 68
One Step at a Time: Finding the Mean 69
Three Characteristics of the Mean 69
Using SPSS to Produce Measures of Central Tendency 73
Choosing a Measure of Central Tendency 75
Applying Statistics 3.1: The Mean, Mode, and Median 76
SUMMARY / 78 • SUMMARY OF FORMULAS / 78 •
GLOSSARY / 78 • PROBLEMS / 78
YOU ARE THE RESEARCHER: THE TYPICAL AMERICAN 85

Chapter 4 Measures of Dispersion 87

The Range (R) and Interquartile Range (Q) 89
Computing the Range and Interquartile Range 90
One Step at a Time: Finding the Interquartile Range (Q) 91
The Standard Deviation and Variance 92
One Step at a Time: Finding the Standard Deviation (s) and the Variance (s^2) 94
Applying Statistics 4.1: The Standard Deviation 95
Computing the Standard Deviation: An Additional Example 95
Visualizing Dispersion: Boxplots 97
Interpreting the Standard Deviation 100
Applying Statistics 4.2: Describing Dispersion 101
Using SPSS to Produce Measures of Dispersion 102
Becoming a Critical Consumer: Central Tendency, Dispersion, and the Shape of Distributions 104
SUMMARY / 106 • SUMMARY OF FORMULAS / 107 •
GLOSSARY / 107 • PROBLEMS / 107
YOU ARE THE RESEARCHER: FOLLOWING UP ON EARLIER PROJECTS 113

Chapter 5 The Normal Curve 118

Properties of the Normal Curve 118
Using the Normal Curve 121
One Step at a Time: Finding Z Scores 122
The Normal Curve Table (Appendix A) 123
Finding the Total Area Above and Below a Score 125
One Step at a Time: Finding Areas Above and Below Positive and Negative Z Scores 127
Finding Areas Between Two Scores 127
One Step at a Time: Finding Areas Between Z Scores 129
Applying Statistics 5.1: Finding Z Scores and Areas Below 129
Applying Statistics 5.2: Finding Z Scores and Areas Between 129
Using the Normal Curve to Estimate Probabilities 130
One Step at a Time: Finding Probabilities 133
Applying Statistics 5.3: Finding Probabilities 134
SUMMARY / 134 • SUMMARY OF FORMULAS / 134 •
GLOSSARY / 135 • PROBLEMS / 135

Part II Inferential Statistics 139

Chapter 6 Introduction to Inferential Statistics, the Sampling Distribution, and Estimation 140

Probability Sampling: Basic Concepts 141
The Sampling Distribution 143
The Sampling Distribution: An Additional Example Using the General Social Survey (GSS) 148
Symbols and Terminology 150
Introduction to Estimation 151
Estimation Selection Criteria: Bias and Efficiency 151
Interval Estimation Procedures 154
Interval Estimation Procedures Using Sample Means (Large Samples) 157
One Step at a Time: Constructing Confidence Intervals for Sample Means by Using Formula 6.2 159
One Step at a Time: Interpreting Confidence Intervals Using Sample Means 159
Applying Statistics 6.1: Estimating a Population Mean 159
Interval Estimation Procedures Using Sample Proportions (Large Samples) 160
One Step at a Time: Constructing Confidence Intervals Using Sample Proportions 161
One Step at a Time: Interpreting Confidence Intervals Using Sample Proportions 161
Applying Statistics 6.2: Estimating Population Proportions 162
Computing Confidence Intervals: A Summary 163
Controlling the Width of Interval Estimates 163
Using SPSS to Construct Confidence Intervals 165
Becoming a Critical Consumer: Public Opinion Polls 167
SUMMARY / 169 • SUMMARY OF FORMULAS / 169 •
GLOSSARY / 169 • PROBLEMS / 170
YOU ARE THE RESEARCHER: ESTIMATING THE CHARACTERISTICS OF THE TYPICAL AMERICAN 173

Chapter 7 Hypothesis Testing I: The One-Sample Case 176

An Overview of Hypothesis Testing 177
The Five-Step Model for Hypothesis Testing 182
One Step at a Time: Completing Step 4 of the Five-Step Model: Computing Z(obtained) 184
One Step at a Time: Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results 185
Choosing a One-Tailed or Two-Tailed Test 186
Selecting an Alpha Level 191
The Student's *t* Distribution 193
Applying Statistics 7.1: Testing a Sample Mean for Significance 194
One Step at a Time: Completing Step 4 of the Five-Step Model: Computing t(Obtained) 197
One Step at a Time: Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results 197
Tests of Hypothesis for Single-Sample Proportions (Large Samples) 198
One Step at a Time: Completing Step 4 of the Five-Step Model: Computing Z(obtained) 200
One Step at a Time: Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results 201
Applying Statistics 7.2: Testing a Sample Proportion for Significance 201
SUMMARY / 202 • SUMMARY OF FORMULAS / 203 • GLOSSARY / 203 • PROBLEMS / 204

Chapter 8 Hypothesis Testing II: The Two-Sample Case 207

The One-Sample Case Versus the Two-Sample Case 208
Hypothesis Testing with Sample Means (Large Samples) 209
One Step at a Time: Completing Step 4 of the Five-Step Model: Computing Z(obtained) 212
One Step at a Time: Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results 212
Applying Statistics 8.1: A Test of Significance for Sample Means 213
Hypothesis Testing with Sample Means (Small Samples) 214
One Step at a Time: Completing Step 4 of the Five-Step Model: Computing t(obtained) 216
One Step at a Time: Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results 217
Using SPSS to Test Sample Means for Significant Differences 217
Hypothesis Testing with Sample Proportions (Large Samples) 219
One Step at a Time: Completing Step 4 of the Five-Step Model: Computing Z(obtained) 221
One Step at a Time: Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results 222

Applying Statistics 8.2: Testing the Significance of the Difference Between Sample Proportions 222
The Limitations of Hypothesis Testing: Significance Versus Importance 224
Becoming a Critical Consumer: When is a difference a difference? 226
SUMMARY / 227 • SUMMARY OF FORMULAS / 228 •
GLOSSARY / 228 • PROBLEMS / 229
YOU ARE THE RESEARCHER: GENDER GAPS AND SUPPORT FOR TRADITIONAL GENDER ROLES 233

Chapter 9 Hypothesis Testing III: The Analysis of Variance 238

The Logic of the Analysis of Variance 239
The Computation of ANOVA 240
An Example of Computing the Analysis of Variance 243
A Test of Significance for ANOVA 244
One Step at a Time: Completing Step 4 of the Five-Step Model: Computing F(obtained) 246
One Step at a Time: Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results 247
Applying Statistics 9.1: The Analysis of Variance 247
An Additional Example of Computing and Testing the Analysis of Variance 249
Using SPSS: The ANOVA Test 251
The Limitations of the Test 253
Becoming a Critical Consumer: Reading the Professional Literature 254
SUMMARY / 255 • SUMMARY OF FORMULAS / 255 •
GLOSSARY / 256 • PROBLEMS / 256
YOU ARE THE RESEARCHER: WHY ARE SOME PEOPLE LIBERAL (OR CONSERVATIVE)? 261

Chapter 10 Hypothesis Testing IV: Chi Square 265

Bivariate Tables 266
The Logic of Chi Square 268
The Computation of Chi Square 269
One Step at a Time: Computing Expected Frequencies (f_e) 271
One Step at a Time: Computing Chi Square 271
The Chi Square Test for Independence 272
Column Percentages 273
One Step at a Time: Computing Column Percentages 274
Applying Statistics 10.1: The Chi Square Test 275
The Chi Square Test: An Additional Example 276
Using SPSS: Crosstabs and Chi Square 278
The Limitations of the Chi Square Test 279
Applying Statistics 10.2: Who Reads Horoscopes? 282
SUMMARY / 282 • SUMMARY OF FORMULAS / 283 •
GLOSSARY / 283 • PROBLEMS / 283
YOU ARE THE RESEARCHER: UNDERSTANDING POLITICAL BELIEFS 287

Part III Bivariate Measures of Association 291

Chapter 11 Bivariate Association for Nominal- and Ordinal-Level Variables 292

Association and Bivariate Tables 293
Three Characteristics of Bivariate Associations 295
Association and Statistical Significance 299
Association and Causation 300
Applying Statistics 11.1: The Reading Bivariate Tables and Avoiding Errors of Interpretation 300
Measures of Association: Introduction 301
Measures of Association for Nominal Variables 302
One Step at a Time: Computing Cramer's V 303
One Step at a Time: Interpreting Cramer's V 304
One Step at a Time: Computing Lambda 306
One Step at a Time: Interpreting Lambda 306
Applying Statistics 11.2: Using Cramer's V and Lambda 307
Measures of Association for Collapsed Ordinal Variables 308
One Step at a Time: Computing Gamma 311
One Step at a Time: Interpreting Gamma 312
Applying Statistics 11.3: Interpreting Direction with Ordinal-Level Variables 312
Using SPSS: Crosstabs and Gamma 312
A Measure of Association for Continuous Ordinal Variables: Spearman's Rho (r_s) 315
One Step at a Time: Computing Spearman's Rho 317
One Step at a Time: Interpreting Spearman's Rho 317
Applying Statistics 11.4: Spearman's Rho 318
Becoming a Critical Consumer: Bivariate Tables in the Professional Literature 319
SUMMARY / 319 • SUMMARY OF FORMULAS / 320 •
GLOSSARY / 320 • PROBLEMS / 321
YOU ARE THE RESEARCHER: UNDERSTANDING POLITICAL BELIEFS, PART II 325

Chapter 12 Association Between Variables Measured at the Interval-Ratio Level 328

Scatterplots and Regression Analysis 329
The Regression Line 335
One Step at a Time: Computing the Slope (b) 339
One Step at a Time: Computing the Y Intercept (a) 340
One Step at a Time: Using the Regression Line to Predict Scores on Y 340
The Correlation Coefficient (Pearson's *r*) 341
One Step at a Time: Computing Pearson's r 341
Interpreting the Correlation Coefficient: r^2 342
One Step at a Time: Interpreting the Strength and Direction of Pearson's r 345
Applying Statistics 12.1: Computing and Interpreting the Regression Coefficients and Pearson's r 346

Using SPSS to Conduct a Regression Analysis and Calculate Pearson's r and r^2 347
The Correlation Matrix 347
Using SPSS to Produce a Correlation Matrix 348
Correlation and Causation 350
Correlation and Statistical Significance 350
Correlation, Regression, Level of Measurement, and Dummy Variables 351
Using SPSS to Produce a Regression Analysis with a Dummy Variable 352
Applying Statistics 12.2: Fertility and Women's Education 353
Becoming a Critical Consumer: Correlation, Causation, and Cancer 354
SUMMARY / 354 • SUMMARY OF FORMULAS / 355 •
GLOSSARY / 355 • PROBLEMS / 356
YOU ARE THE RESEARCHER: WHO WATCHES TV?
WHO SUCCEEDS IN LIFE? 360

Part IV Multivariate Techniques 363

Chapter 13 Partial Correlation and Multiple Regression and Correlation 364

Partial Correlation 365
One Step at a Time: Computing the Partial Correlation Coefficient 370
One Step at a Time: Interpreting the Partial Correlation Coefficient 371
Applying Statistics 13.1: Using SPSS to Determine if the Relationship Between Fertility and Women's Education Is Affected by Affluence 371
Multiple Regression: Predicting the Dependent Variable 372
One Step at a Time: Computing Partial Slopes 374
One Step at a Time: Computing the Y Intercept 375
One Step at a Time: Using the Multiple Regression Line to Predict Scores on Y 376
Multiple Regression: Assessing the Effects of the Independent Variables 376
One Step at a Time: Computing Beta-Weights (b^*) 378
One Step at a Time: Interpreting Beta-Weights (b^*) 378
Multiple Correlation 379
One Step at a Time: Computing the Coefficient of Multiple Determination (R^2) 380
One Step at a Time: Interpreting R^2 380
Using SPSS to Conduct a Regression Analysis 380
Applying Statistics 13.2: Multiple Regression and Correlation 382
The Limitations of Multiple Regression and Correlation 384
Becoming a Critical Consumer: Reading the Professional Literature 385
SUMMARY / 386 • SUMMARY OF FORMULAS / 386 •
GLOSSARY / 387 • PROBLEMS / 387
YOU ARE THE RESEARCHER: A MULTIVARIATE ANALYSIS OF TV VIEWING AND SUCCESS 392

Appendix A Area Under the Normal Curve 395

Appendix B Distribution of t 399

Appendix C Distribution of Chi Square 400

Appendix D Distribution of F 401

Appendix E Using Statistics: Ideas for Research Projects 403

Appendix F An Introduction to SPSS 409

Appendix G Codebooks for All Datasets 417

Answers to Odd-Numbered End-of-Chapter Problems 429

Glossary 441

Index 447

Preface

Statistics are part of the everyday language of sociology and other social sciences (including political science, social work, public administration, criminal justice, urban studies, and gerontology). These disciplines are research-based and routinely use statistics to express knowledge and to discuss theory and research. To join the conversations being conducted in these disciplines, you must be literate in the vocabulary of research, data analysis, and scientific thinking. Knowledge of statistics will enable you to understand the professional research literature, conduct quantitative research yourself, contribute to the growing body of social science knowledge, and reach your full potential as a social scientist.

Although essential, learning (and teaching) statistics can be a challenge. Students in social science statistics courses typically have a wide range of mathematical backgrounds and an equally diverse set of career goals. They are often puzzled about the relevance of statistics for them, and, not infrequently, there is some math anxiety to deal with. This text introduces statistical analysis for the social sciences while addressing these realities.

The text makes minimal assumptions about mathematical background (the ability to read a simple formula is sufficient preparation for virtually all of the material in the text), and a variety of special features help students analyze data successfully. The text has been written especially for sociology and social work programs but is sufficiently flexible to be used in any program with a social science base.

The text is written at an intermediate level and its intent is to show the relevance and value of statistics for the social sciences. I emphasize interpretation and understanding statistics in the context of social science research, but I have not sacrificed comprehensive coverage or statistical correctness. Mathematical explanations are kept at an elementary level, as is appropriate in a first exposure to social statistics. For example, I do not treat formal probability theory per se in the text.[1] Rather, the background necessary for an understanding of inferential statistics is introduced, informally and intuitively, in Chapters 5 and 6 while considering the concepts of the normal curve and the sampling distribution.

The text does not claim that statistics are "fun" or that the material can be mastered without considerable effort. At the same time, students are not overwhelmed with abstract proofs, formula derivations, or mathematical theory, which can needlessly frustrate the learning experience at this level.

[1]A presentation of probability is available at the website for this text for those who are interested.

Goals of the Text

The primary goal of this text is to develop basic statistical literacy. The statistically literate person understands and appreciates the role of statistics in the research process, is competent to perform basic calculations, and can read and appreciate the professional research literature in his or her field as well as any research reports he or she may encounter outside academia.

The goal of achieving basic statistical literacy has not changed since the first edition of this text. However, in recognition of the fact that "mere computation" has become less of a challenge in this high-tech age, this edition continues to increase the stress on interpretation and computer applications while deemphasizing computation. This will be apparent in several ways:

- In recognition of the fact that modern technology has rendered hand calculation increasingly obsolete, the end-of-chapter problems feature smaller, easier-to-handle datasets, although some more challenging problems are also included.
- A section called "Using SPSS" has been added to most chapters to demonstrate how to use a computerized statistical package to produce the statistics covered in that chapter.
- The end-of-chapter problems now include SPSS-based exercises, and research projects using SPSS are included at the end of almost all chapters.
- To accommodate the increased use of SPSS, several new datasets have been added to the text and the General Social Survey dataset has been updated to 2012.

The three aspects of basic statistical literacy provide a framework for discussing the additional features of this text.

1. **An Appreciation of Statistics.** A statistically literate person understands the relevance of statistics for social research, can analyze and interpret the meaning of a statistical test, and can select an appropriate statistic for a given purpose and a given set of data. This textbook develops these skills, within the constraints imposed by the introductory nature of the course, in several ways.
   - *The relevance of statistics.* Chapter 1 includes a discussion of the role of statistics in social research and stresses their usefulness as ways of analyzing and manipulating data and answering research questions. Throughout the text, each example problem is framed in the context of a research situation. A question is posed and then, with the aid of a statistic, answered. This central theme of usefulness is further reinforced by a series of "Applying Statistics" boxes, each of which illustrates some specific way statistics can be used to answer questions, and by the "Using Statistics" feature that opens every chapter.

     End-of-chapter problems are labeled by the social science discipline from which they are drawn: ***SOC*** for sociology, ***SW*** for social work, ***PS*** for political science, ***CJ*** for criminal justice, ***PA*** for public administration, and ***GER*** for gerontology. Identifying problems with specific disciplines

allows students to more easily see the relevance of statistics to their own academic interests. (Not incidentally, they will also see that the disciplines have a large subject matter in common.)

Also, a series of boxed features entitled "Statistics in Everyday Life" appear in each chapter and highlight the relevance of statistics in the real world and provide examples of everyday applications.

- *Interpreting statistics.* For most students, interpretation—saying what statistics mean—is a big challenge. The ability to interpret statistics can be developed only by exposure and experience. To provide exposure, I have been careful, in the example problems, to express the meaning of the statistic in terms of the original research question. To provide experience, the end-of-chapter problems call for an interpretation of the statistic calculated. To provide examples, many of the answers to odd-numbered computational problems in the back of the text are expressed in words as well as numbers.
- *Using statistics: Ideas for research projects*. Appendix E offers ideas for independent data-analysis projects for students. The projects require students to use SPSS to analyze a dataset. They can be assigned at intervals throughout the semester or at the end of the course. Each project provides an opportunity for students to practice and apply their statistical skills and, above all, to exercise their ability to understand and interpret the statistics they produce.

2. **Computational Competence.** Students should emerge from their first course in statistics with the ability to perform elementary forms of data analysis. While computers have made computation less of an issue today, computation is inseparable from statistics, so I have included a number of features to help students cope with these mathematical challenges.
   - *"One Step at a Time"* boxes for each statistic break down computation into individual steps for maximum clarity and ease.
   - *Extensive problem sets* are provided at the end of each chapter. For the most part, these problems use fictitious data and are designed for ease of computation.
   - *Solutions* to odd-numbered computational problems are provided so that students may check their answers.
   - *SPSS* gives students access to the computational power of the computer. This is explained in more detail later.

3. **The Ability to Read the Professional Social Science Literature.** The statistically literate person can comprehend and critically appreciate research reports written by others. The development of this skill is a particular problem at the introductory level because (1) the vocabulary of professional researchers is so much more concise than the language of the textbook, and (2) the statistics featured in the literature are more advanced than those covered at the introductory level. This text helps to bridge this gap by
   - Always expressing the meaning of each statistic in terms of answering a social science research question.

- Providing a series of boxed inserts called "Becoming a Critical Consumer" that help students decipher the statistics they are likely to encounter in everyday life as well as in the professional literature.

Additional Features

A number of other features make the text more meaningful for students and more useful for instructors.

- *Readability and clarity.* The writing style is informal and accessible to students without ignoring the traditional vocabulary of statistics. Problems and examples have been written to maximize student interest and to focus on issues of concern and significance. For the more difficult material (such as hypothesis testing), students are first walked through an example problem before being confronted by formal terminology and concepts. Each chapter ends with a summary of major points and formulas and a glossary of important concepts. A list of frequently used formulas inside the covers and a glossary can be used for quick reference.
- *Organization and coverage.* The text is divided into four parts, with most of the coverage devoted to univariate descriptive statistics, inferential statistics, and bivariate measures of association. The distinction between description and inference is introduced in the first chapter and maintained throughout the text. In selecting statistics for inclusion, I have tried to strike a balance between the essential concepts with which students must be familiar and the amount of material students can reasonably be expected to learn in their first (and perhaps only) statistics course, all the while bearing in mind that different instructors will naturally wish to stress different aspects of the subject. Thus, the text covers a full gamut of the usual statistics, with each chapter broken into subsections so that instructors may choose the particular statistics they wish to include.
- *Learning objectives.* Learning objectives are stated at the beginning of each chapter. These are intended to serve as "study guides" and to help students identify and focus on the most important material.
- *Using Statistics.* At the beginning of each chapter, some applications of the statistics to be introduced are presented to give students a context for appreciating the material and some further examples of the usefulness of statistics.
- *Review of mathematical skills.* A comprehensive review of all of the mathematical skills that will be used in this text is included as a Prologue. Students who are inexperienced or out of practice with mathematics can study this review early in the course and/or refer to it as needed. A self-test is included so that students can check their level of preparation for the course.
- *Statistical techniques and end-of-chapter problems are explicitly linked.* After a technique is introduced, students are directed to specific problems for practice and review. The "how-to-do-it" aspects of calculation are immediately and clearly reinforced.
- *End-of-chapter problems are organized progressively.* Simpler problems with small datasets are presented first. Often, explicit instructions or hints accompany

the first several problems in a set. The problems gradually become more challenging and require more decision making by the student (e.g., choosing the most appropriate statistic for a certain situation). Thus, each problem set develops problem-solving abilities gradually and progressively.

- *Computer applications.* This text integrates SPSS, the leading social science statistics package, to help students take advantage of the power of the computer. Appendix F provides an introduction to SPSS, and demonstrations are integrated into the chapters. SPSS-based problems are included at the end of chapters, and research projects using SPSS are presented in the "You Are the Researcher" feature.
- *Realistic, up-to-date data.* The databases for computer applications in the text include a shortened version of the 2012 General Social Survey, a dataset that includes census and crime data for the 50 states, and a dataset that includes demographic data for 99 nations. These databases will give students the opportunity to practice their statistical skills on "real-life" data. All databases are described in Appendix G.

Additional Course Design Resources

- *Online PowerPoint® Slides.* A revised series of PowerPoint slides allows instructors to deliver class lectures and presentations discussing chapter-by-chapter content.
- *Online Instructor's Manual/Testbank.* The Instructor's Manual includes chapter summaries, a test item file of multiple-choice questions, answers to even-numbered computational problems, and step-by-step solutions to selected problems. In addition, the Instructor's Manual includes cumulative exercises (with answers) that can be used for testing purposes. To access these instructor resources, please log in to your account at **https://login.cengage.com**.
- *Aplia*™ is an online interactive learning solution that can be assigned as part of the course. Aplia integrates a variety of media and tools such as video, tutorials, practice tests, and an interactive e-book, and provides students with detailed, immediate feedback on every question. For more information about how to use Aplia in your course, please work with your local Cengage Learning Consultant.

Changes to the Fourth Edition

The following are the most important changes to this edition of *Essentials*:

- SPSS has been moved to a more central place in the text:
  - Almost all chapters have new sections ("Using SPSS") that illustrate how to produce the statistics covered in the chapter.
  - SPSS problems have been added to the end-of-chapter problems throughout the text. In some chapters (e.g., Chapters 12 and 13), the SPSS problems replace problems using hand calculators.

  - For statistics that require complex computation—such as Pearson's *r* (Chapter 12) and partial correlation, multiple correlation, and regression (Chapter 13)—explanations and examples are now SPSS-based.
- The datasets used in the text have been expanded and updated. The datasets are used throughout the text in the new "Using SPSS" sections, in the end-of-chapter problems, and in the "You Are the Researcher" projects at the end of most chapters. The datasets are available for downloading at the website for this text: **www.cengagebrain.com**; they include
  - A General Social Survey (GSS) dataset (*GSS2012.sav*), which has been updated to 2012.
  - A dataset that includes census and crime data on the 50 states (*States.sav*).
  - A dataset that includes mostly demographic data for 99 nations (*Intl-POP.sav*).
  - A fourth dataset (*CrimeTrends84-10.sav*) is used only for the graphing exercises in Chapter 2.
- Former Chapters 11 and 12 have been combined into a single chapter (Chapter 11, entitled "Bivariate Association for Nominal- and Ordinal-Level Variables"). This new chapter de-emphasizes phi and the mechanics of computation for gamma but still fully treats the analysis of association for variables organized in bivariate tables.
- Chapter 2 has been reorganized and now begins with frequency distributions.
- Boxplots have been added to Chapter 4.
- All chapters now begin with a "Using Statistics" box that cites examples of how the statistics presented in the chapter can be applied to social research and to everyday life.
- Most of the "Statistics in Everyday Life" boxes have been updated or changed.
- The "Becoming a Critical Consumer" inserts have been updated and many have been shortened or broken into separate boxes
- The datasets used for examples, in the boxed features, and in the end-of-chapter problems have been updated.
- Titles have been added to all boxed features to clarify content and purpose.

Numerous other changes have been made throughout the text, most of them minor. All are intended to clarify explanations and make the material more accessible to students. As with previous editions, my goal is to offer a comprehensive, flexible, and student-oriented book that will provide a challenging first exposure to social statistics.

Acknowledgments

This text, in one form or another, has been in development for 30 years. An enormous number of people have made contributions, both great and small, to this project and, at the risk of inadvertently omitting someone, I am bound to at least attempt to acknowledge my many debts.

Much of whatever integrity and quality this book has is a direct result of the very thorough (and often highly critical) reviews that have been conducted over the years. I am consistently impressed by the sensitivity of my colleagues to the needs of the students. Whatever failings are contained in the text are, of course, my responsibility and are probably the result of my occasional decisions not to follow the advice of my colleagues.

I would like to thank the instructors who made statistics understandable to me (Professors Satoshi Ito, Noelie Herzog, and Ed Erikson) and all of my colleagues at Christopher Newport University for their support and encouragement (especially Professors F. Samuel Bauer, Stephanie Byrd, Cheryl Chambers, Robert Durel, Marcus Griffin, Mai Lan Gustafsson, Kai Heiddemann, Ruth Kernodle, Michael Lewis, Marion Manton, Eileen O'Brien, Lea Pellet, Eduardo Perez, Virginia Purtle, Andrea Timmer, and Linda Waldron). Also, I thank all of my students for their patience and thoughtful feedback, and I am grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Longman Group Ltd., London, for permission to reprint Appendices B, C, and D from their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974).

Finally, I want to acknowledge the support of my family and rededicate this work to them. I have the extreme good fortune to be a member of an extended family that is remarkable in many ways and that continues to increase in size. Although I cannot list everyone, I would especially like to thank the older generation (my mother, Alice T. Healey, may she rest in peace), my wife Patricia A. Healey, the next generation (my sons, Kevin and Christopher, my daughters-in-law, Jennifer and Jessica, my step-son Christopher Schroen, and step-daughters Kate Cowell and her husband Matt and Jennifer Schroen), and the youngest generation (Benjamin, Caroline, and Isabelle Healey and Abigail Cowell).

Prologue: Basic Mathematics Review

You will probably be relieved to hear that this text, your first exposure to statistics for social science research, does not stress computation per se. While you will encounter many numbers to work with and numerous formulas to solve, the major emphasis will be on understanding the role of statistics in research and the logic we use to answer research questions empirically. You will also find that, in this text, the example problems and many of the homework problems have been intentionally simplified so that the computations will not unduly impede the task of understanding the statistics themselves.

On the other hand, you may regret to learn that there is, inevitably, some arithmetic that you simply cannot avoid if you want to master this material. It is likely that some of you haven't had any math in a long time, others have convinced themselves that they just cannot do math under any circumstances, and still others are just rusty and out of practice. All of you will find that mathematical operations that might seem complex and intimidating at first glance can be broken down into simple steps. If you have forgotten how to cope with some of these steps or are unfamiliar with these operations, this prologue is designed to ease you into the skills you will need to do all of the computations in this textbook. Also, you can use this section for review whenever you feel uncomfortable with the mathematics in the chapters to come.

Calculators and Computers

A calculator is a virtual necessity for this text. Even the simplest, least expensive model will save you time and effort and is definitely worth the investment. However, I recommend that you consider investing in a more sophisticated calculator with memories and preprogrammed functions, especially the statistical models that can compute means and standard deviations automatically. Calculators with these capabilities are available for around $20.00 to $30.00 and will almost certainly be worth the small effort it will take to learn to use them.

In the same vein, there are several computerized statistical packages (or **statpaks**) commonly available on college campuses that can further enhance your statistical and research capabilities. The most widely used of these is the Statistical Package for the Social Sciences (**SPSS**). Statistical packages like SPSS are many times more powerful than even the most sophisticated handheld calculators, and it will be well worth your time to learn how to use them because they will eventually save you time and effort.

SPSS is introduced in Appendix F of this text and is integrated into almost all the chapters. There are demonstrations that show you, step by step, how to use the program to generate the statistics covered in the chapter and end-of-chapter problems that require you to apply the program. Furthermore, the "You Are the Researcher" feature at the end of most chapters gives you the opportunity to use SPSS in some simplified social research projects.

There are many other programs that can help you calculate statistics with a minimum of effort and time. Even spreadsheet programs such as Microsoft® Excel, which is included in many versions of Microsoft Office, have some statistical capabilities. You should be aware that all of these programs (other than the simplest calculators) will require some effort to learn, but the rewards will be worth the effort.

In summary, you should find a way at the beginning of this course—with a calculator, a statpak, or both—to minimize the tedium of mere computing. This will permit you to devote maximum effort to the truly important goal of increasing your understanding of the meaning of statistics in particular and social research in general.

Variables and Symbols

Statistics are a set of techniques by which we can describe, analyze, and manipulate variables. A **variable** is a trait that can change value from case to case or from time to time. Examples of variables include height, weight, level of prejudice, and political party preference. The possible values or scores associated with a given variable might be numerous (for example, income) or relatively few (for example, gender). I will often use symbols, usually the letter X, to refer to variables in general or to a specific variable.

Sometimes we will need to refer to a specific value or set of values of a variable. This is usually done with the aid of subscripts. So the symbol X_1 (read "X-sub-one") would refer to the first score in a set of scores, X_2 ("X-sub-two") to the second score, and so forth. Also, we will use the subscript i to refer to all the scores in a set. Thus, the symbol X_i ("X-sub-i") refers to all of the scores associated with a given variable (for example, the test grades of a particular class).

Operations

You are all familiar with the four basic mathematical operations of addition, subtraction, multiplication, and division and the standard symbols ($+$, $-$, $\times$, $\div$) used to denote them. I should remind you that multiplication and division can be symbolized in a variety of ways. For example, the operation of multiplying some number a by some number b may be symbolized in (at least) six different ways:

$$a \times b$$
$$a \cdot b$$
$$a * b$$

$$ab$$
$$a(b)$$
$$(a)(b)$$

In this text, we will commonly use the "adjacent symbols" format (that is, ab), the conventional times sign ($\times$), or adjacent parentheses to indicate multiplication. On most calculators and computers, the asterisk (*) is the symbol for multiplication.

The operation of division can also be expressed in several different ways. In this text, in addition to the standard symbol for division, we will use either of these two methods:

$$a/b \text{ or } \frac{a}{b}$$

Several formulas require us to find the square of a number. To do this, multiply the number by itself. This operation is symbolized as X^2 (read "X squared"), which is the same thing as $(X)(X)$. If X has a value of 4, then

$$X^2 = (X)(X) = (4)(4) = 16$$

or we could say that "4 squared is 16."

The square root of a number is the value that, when multiplied by itself, results in the original number. So the square root of 16 is 4 because (4)(4) is 16. The operation of finding the square root of a number is symbolized as

$$\sqrt{X}$$

A final operation with which you should be familiar is summation, or the addition of the scores associated with a particular variable. When a formula requires the addition of a series of scores, this operation is usually symbolized as ΣX_i. "Σ" is uppercase Greek letter sigma and stands for "the summation of." So the combination of symbols ΣX_i means "the summation of all the scores" and directs us to add all the scores for that variable. If four people had family sizes of 2, 4, 5, and 7, then the summation of these four scores for this variable could be symbolized as

$$\Sigma X_i = 2 + 4 + 5 + 7 = 18$$

The symbol Σ is an operator, just like the $+$ and $\times$ signs. It directs us to add all of the scores on the variable indicated by the X symbol.

There are two other common uses of the summation sign, and, unfortunately, the symbols denoting these uses are not, at first glance, sharply different from each other or from the symbol used earlier. Some careful attention to these various meanings should minimize the confusion.

The first set of symbols is ΣX^2, which means "the sum of the squared scores." This quantity is found by *first* squaring each of the scores and *then* adding the squared scores together. A second common set of symbols will be $(\Sigma X_i)^2$, which means "the sum of the scores, squared." This quantity is found by *first* summing the scores and *then* squaring the total.

These distinctions might be confusing at first, so let's use an example to help clarify the situation. Suppose we had a set of three scores: 10, 12, and 13. So

$$X_i = 10, 12, 13$$

The sum of these scores is

$$\Sigma X_i = 10 + 12 + 13 = 35$$

The sum of the squared scores would be

$$\Sigma X^2 = (10)^2 + (12)^2 + (13)^2 = 100 + 144 + 169 = 413$$

Take careful note of the order of operations here. First the scores are squared one at a time, and then the squared scores are added. This is a completely different operation from squaring the sum of the scores:

$$(\Sigma X_i)^2 = (10 + 12 + 13)^2 = (35)^2 = 1225$$

To find this quantity, first the scores are summed and then the total of all the scores is squared. The squared sum of the scores (1225) is *not* the same as the sum of the squared scores (413).

In summary, the operations associated with each set of symbols can be summarized as follows:

Symbols	Operations
ΣX_i	Add the scores.
ΣX_i^2	First square the scores and then add the squared scores.
$(\Sigma X_i)^2$	First add the scores and then square the total.

Operations with Negative Numbers

A number can be either positive (if it is preceded by a + sign or by no sign at all) or negative (if it is preceded by a − sign). Positive numbers are greater than 0, and negative numbers are less than 0. It is very important to keep track of signs because they will affect the outcome of virtually every mathematical operation. This section briefly summarizes the relevant rules for dealing with negative numbers.

First, adding a negative number is the same as subtraction. For example,

$$3 + (-1) = 3 - 1 = 2$$

Second, subtraction changes the sign of a negative number:

$$3 - (-1) = 3 + 1 = 4$$

Note the importance of keeping track of signs here. If you neglected to change the sign of the negative number in the second expression, you would get the wrong answer.

For multiplication and division, you need to be aware of various combinations of negative and positive numbers. Ignoring the case of all positive numbers, this

leaves several possible combinations. A negative number times a positive number results in a negative value:

$$(-3)(4) = -12$$
$$(3)(-4) = -12$$

A negative number multiplied by a negative number is always positive:

$$(-3)(-4) = 12$$

Division follows the same patterns. If there is a negative number in the calculations, the answer will be negative. If both numbers are negative, the answer will be positive. So

$$\frac{-4}{2} = -2$$

and

$$\frac{4}{-2} = -2$$

but

$$\frac{-4}{-2} = 2$$

Negative numbers do not have square roots, because multiplying a number by itself cannot result in a negative value. Squaring a negative number always results in a positive value (see the multiplication rules earlier).

Accuracy and Rounding Off

A possible source of confusion in computation involves accuracy and rounding off. People work at different levels of precision and, for this reason alone, may arrive at different answers to problems. This is important because our answers can be at least slightly different if you work at one level of precision and I (or your instructor or your study partner) work at another. You may sometimes think you've gotten the wrong answer when all you've really done is round off at a different place in the calculations or in a different way.

There are two issues here: when to round off and how to round off. My practice is to work in as much accuracy as my calculator or statistics package will allow and then round off to two places of accuracy (two places beyond, or to the right of, the decimal point) only at the very end. If a set of calculations is lengthy and requires the reporting of intermediate sums or subtotals, I will round the subtotals off to two places as I go.

In terms of how to round off, begin by looking at the digit immediately to the right of the last digit you want to retain. If you want to round off to 100ths (two places beyond the decimal point), look at the digit in the 1000ths place (three places beyond the decimal point). If that digit is 5 or more, round up. For

example, 23.346 would round off to 23.35. If the digit to the right is less than 5, round down. So, 23.343 would become 23.34.

Let's look at an additional example of how to follow these rules of rounding. If you are calculating the mean value of a set of test scores and your calculator shows a final value of 83.459067 and you want to round off to two places, look at the digit three places beyond the decimal point. In this case the value is 9 (greater than 5), so we would round the second digit beyond the decimal point up and report the mean as 83.46. If the value had been 83.453067, we would have reported our final answer as 83.45.

Formulas, Complex Operations, and the Order of Operations

A mathematical formula is a set of directions, stated in general symbols, for calculating a particular statistic. To "solve a formula," you replace the symbols with the proper values and then perform a series of calculations. Even the most complex formula can be simplified by breaking the operations down into smaller steps.

Working through the steps requires some knowledge of general procedure and the rules of precedence of mathematical operations. This is because the order in which you perform calculations may affect your final answer. Consider the following expression:

$$2 + 3(4)$$

If you add first, you will evaluate the expression as

$$5(4) = 20$$

but if you multiply first, the expression becomes

$$2 + 12 = 14$$

Obviously, it is crucial to complete the steps of a calculation in the correct order.

The basic rules of precedence are to find all squares and square roots first, then do all multiplication and division, and finally complete all addition and subtraction. So the expression

$$8 + 2 \times 2^2/2$$

would be evaluated as

$$8 + 2 \times \frac{4}{2} = 8 + \frac{8}{2} = 8 + 4 = 12$$

The rules of precedence may be overridden by parentheses. Solve all expressions within parentheses before applying the rules stated earlier. For most formulas in this text, the order of calculations will be controlled by the parentheses.

Consider the following expression:

$$(8 + 2) - \frac{4}{(5 - 1)}$$

Resolving the parenthetical expressions first, we would have

$$(8 + 2) - \frac{4}{(5 - 1)} = (10) - \frac{4}{4} = 10 - 1 = 9$$

A final operation you will encounter in some formulas in this text involves denominators of fractions that themselves contain fractions. In this situation, solve the fraction in the denominator first and then complete the division. For example,

$$\frac{15 - 9}{6/2}$$

would become

$$\frac{15 - 9}{6/2} = \frac{6}{3} = 2$$

When you are confronted with complex expressions such as these, don't be intimidated. If you're patient with yourself and work through them step by step, beginning with the parenthetical expressions, even the most imposing formulas can be managed.

Exercises

You can use the following problems as a self-test on the material presented in this review. If you can handle these problems, you're ready to do all of the arithmetic in this text. If you have difficulty with any of these problems, please review the appropriate section of this prologue. You might also want to use this section as an opportunity to become more familiar with your calculator. The answers are given immediately following these exercises, along with commentary and some reminders.

1. Complete each of the following:

a. $17 \times 3 =$
b. $17(3) =$
c. $(17)(3) =$
d. $17/3 =$
e. $(42)^2 =$
f. $\sqrt{113} =$

2. For the set of scores (X_i) of 50, 55, 60, 65, and 70, evaluate each of the following expressions:

$\Sigma X_i =$
$\Sigma X_i^2 =$
$(\Sigma X_i)^2 =$

3. Complete each of the following:

a. $17 + (-3) + (4) + (-2) =$
b. $(-27)(54) =$
c. $(-14)(-100) =$
d. $-34/(-2) =$
e. $322/(-11) =$
f. $\sqrt{-2} =$
g. $(-17)^2 =$

4. Round off each of the following to two places beyond the decimal point:

a. 17.17532
b. 43.119
c. 1076.77337
d. 32.4641152301
e. 32.4751152301

5. Evaluate each of the following:

a. $(3 + 7)/10 =$
b. $3 + 7/10 =$
c. $\dfrac{(4 - 3) + (7 + 2)(3)}{(4 + 5)(10)} =$
d. $\dfrac{22 + 44}{15/3} =$

Answers to Exercises

1. **a.** 51 **b.** 51 **c.** 51
(The obvious purpose of these first three problems is to remind you that there are several different ways of expressing multiplication.)
d. 5.67 (Note the rounding off.) **e.** 1764
f. 10.63

2. The first expression is "the sum of the scores," so this operation would be

$$\Sigma X_i = 50 + 55 + 60 + 65 + 70 = 300$$

The second expression is the "sum of the squared scores." So

$$\Sigma X_i^2 = (50)^2 + (55)^2 + (60)^2 + (65)^2 + (70)^2$$
$$\Sigma X_i^2 = 2500 + 3025 + 3600 + 4225 + 4900$$
$$\Sigma X_i^2 = 18{,}250$$

The third expression is "the sum of the scores, squared":

$$(\Sigma X_i)^2 = (50 + 55 + 60 + 65 + 70)^2$$
$$(\Sigma X_i)^2 = (300)^2$$
$$(\Sigma X_i)^2 = 90{,}000$$

Remember that ΣX_i^2 and $(\Sigma X_i)^2$ are two completely different expressions with very different values.

3. **a.** 16
b. −1458
c. 1400
d. 17
e. −29.27
f. Your calculator probably gave you some sort of error message for this problem, because negative numbers do not have square roots.
g. 289

4. **a.** 17.18
b. 43.12
c. 1076.77
d. 32.46
e. 32.48

5. **a.** 1
b. 3.7 (Note the importance of parentheses.)
c. 0.31
d. 13.2