

11th edition

# Statistics

A Tool For Social Research  
& Data Analysis

Joseph F. Healey



# STATISTICS

## A Tool for Social Research and Data Analysis

Eleventh Edition

**Joseph F. Healey**

*Christopher Newport University*

**Christopher Donoghue**

*Montclair State University*



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

Copyright 2021 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. WCN 02-200-322

Copyright 2021 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. The publisher reserves the right to remove content from this title at any time if subsequent rights restrictions require it. For valuable information on pricing, previous editions, changes to current editions, and alternate formats, please visit [www.cengage.com/highered](http://www.cengage.com/highered) to search by ISBN#, author, title, or keyword for materials in your areas of interest.

Important Notice: Media content referenced within the product description or the product text may not be available in the eBook version.

***Statistics: A Tool for Social Research  
and Data Analysis, Eleventh Edition***

Joseph F. Healey, Christopher Donoghue

Product Director: Marta Lee-Perriard

Product Manager: Libby Beiting-Lipps

Content Developer: Sarah Kaubisch

Content Development Manager:  
Trudy Brown

Product Assistant: Allison Balchunas

Content Project Manager: Rita Jaramillo

Digital Content Specialist: Adam Hannah

Customer Account Manager,  
Manufacturing: Karen Hunt

Intellectual Property Manager:  
Betsy Hathaway

Intellectual Property Analyst:  
Deanna Ettinger

Production Service/Composition:  
MPS Limited

© 2021, 2015 Cengage Learning, Inc.

WCN: 02-300

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced or distributed in any form or by any means, except as permitted by U.S. copyright law, without the prior written permission of the copyright owner.

For product information and technology assistance, contact us at  
**Cengage Customer & Sales Support, 1-800-354-9706.**

For permission to use material from this text or product,  
submit all requests online at **[www.cengage.com/permissions](http://www.cengage.com/permissions).**

Further permissions questions can be e-mailed to  
**[permissionrequest@cengage.com](mailto:permissionrequest@cengage.com).**

Library of Congress Control Number: 2020902062

Student Edition:

ISBN: 978-0-357-37107-7

Loose-leaf Edition:

ISBN: 978-0-357-37109-1

**Cengage**

200 Pier 4 Boulevard  
Boston, MA 02210  
USA

Cengage is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **[www.cengage.com](http://www.cengage.com)**.

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

To learn more about Cengage platforms and services, register or access your online learning solution, or purchase materials for your course, visit **[www.cengage.com](http://www.cengage.com)**.

Printed in the United States of America  
Print Number: 01      Print Year: 2020

# Brief Contents

	Preface	xv
	Prologue: Basic Mathematics Review	xxiii
Chapter 1	Introduction	1
<b>Part I</b>	<b>Descriptive Statistics</b>	<b>25</b>
Chapter 2	Basic Descriptive Statistics: Tables, Percentages, Ratios and Rates, and Graphs	26
Chapter 3	Measures of Central Tendency	68
Chapter 4	Measures of Variation	97
Chapter 5	The Normal Curve	129
<b>Part II</b>	<b>Inferential Statistics</b>	<b>153</b>
Chapter 6	Introduction to Inferential Statistics: Sampling and the Sampling Distribution	154
Chapter 7	Estimation Procedures	173
Chapter 8	Hypothesis Testing I: The One-Sample Case	199
Chapter 9	Hypothesis Testing II: The Two-Sample Case	229
Chapter 10	Hypothesis Testing III: The Analysis of Variance	260
Chapter 11	Hypothesis Testing IV: Chi Square	289
<b>Part III</b>	<b>Bivariate Measures of Association</b>	<b>321</b>
Chapter 12	Bivariate Association for Nominal- and Ordinal-Level Variables	322
Chapter 13	Association Between Variables Measured at the Interval-Ratio Level	357
<b>Part IV</b>	<b>Multivariate Techniques</b>	<b>397</b>
Chapter 14	Elaborating Bivariate Tables	398
Chapter 15	Partial Correlation and Multiple Regression and Correlation	424
Appendix A	Area Under the Normal Curve	461
Appendix B	Distribution of $t$	465
Appendix C	Distribution of Chi Square	467
Appendix D	Distribution of $F$	469
Appendix E	Using Statistics: Ideas for Research Projects	471
Appendix F	An Introduction to SPSS	477

Appendix G	Codebooks for All Data Sets	483
Answers to Odd-Numbered End-of-Chapter Problems and Cumulative Exercises		497
Glossary		513
Index		519

# Detailed Contents

Preface	xv
Prologue: Basic Mathematics Review	xxiii
<b>Chapter 1 Introduction</b>	<b>1</b>
Why Study Statistics?	1
The Role of Statistics in Scientific Inquiry	2
<i>Statistics in Everyday Life: Push Polls</i>	5
The Goals of This Text	6
Descriptive and Inferential Statistics	7
<i>Statistics in Everyday Life: Using Descriptive Statistics</i>	8
<i>Statistics in Everyday Life: Using Inferential Statistics</i>	9
Discrete and Continuous Variables	10
Level of Measurement	10
<i>Statistics in Everyday Life: Changes in Socioeconomic Status in the United States</i>	13
<i>One Step at a Time: Determining the Level of Measurement of a Variable</i>	15
<i>Reading Statistics: An Introduction</i>	16
SUMMARY / 19 • GLOSSARY / 19 • PROBLEMS / 20	
YOU ARE THE RESEARCHER: INTRODUCTION	22
<b>Part I Descriptive Statistics</b>	<b>25</b>
<b>Chapter 2 Basic Descriptive Statistics: Tables, Percentages, Ratios and Rates, and Graphs</b>	<b>26</b>
Frequency Distributions for Nominal-Level Variables	27
Enhancing Clarity: Percentages and Proportions	28
<i>One Step at a Time: Finding Percentages and Proportions</i>	31
<i>Social Research and Data Analysis 2.1: Using Proportions and Percentages in Practice</i>	31
<i>Social Research and Data Analysis 2.2: Communicating with Statistics</i>	32
<i>Statistics in Everyday Life</i>	32
Frequency Distributions for Ordinal-Level Variables	32
Frequency Distributions for Interval-Ratio-Level Variables	33
<i>One Step at a Time: Finding Midpoints</i>	36
<i>One Step at a Time: Finding Real Limits</i>	36

<i>One Step at a Time: Adding Cumulative Frequency and Percentage Columns to Frequency Distributions</i>	37
<i>One Step at a Time: Constructing Frequency Distributions for Interval-Ratio Variables</i>	40
<i>Social Research and Data Analysis 2.3: Frequency Distributions</i>	41
<i>Using SPSS: Frequency Distributions</i>	41
Ratios, Rates, and Percentage Change	43
<i>Social Research and Data Analysis 2.4: Ratios</i>	44
<i>Social Research and Data Analysis 2.5: Rates</i>	45
<i>Social Research and Data Analysis 2.6: Percentage Change</i>	47
<i>One Step at a Time: Finding Ratios, Rates, and Percent Change</i>	47
<i>Statistics in Everyday Life: Traffic Fatalities</i>	48
Using Graphs to Present Data	48
<i>Using SPSS: Graphs</i>	53
Interpreting Statistics: Changing Patterns of American Family Life	55
<i>Reading Statistics: Substance Abuse Among College Students</i>	58
SUMMARY / 59 • SUMMARY OF FORMULAS / 60 • GLOSSARY / 60	
• PROBLEMS / 60	
YOU ARE THE RESEARCHER: IS THERE A “CULTURE WAR” IN THE UNITED STATES?	65

### Chapter 3 Measures of Central Tendency

68

The Mode	69
<i>Statistics in Everyday Life: Communication Between Dating Teenagers</i>	70
The Median	71
<i>One Step at a Time: Finding the Median</i>	72
Other Measures of Position: Percentiles, Deciles, and Quartiles	72
<i>Social Research and Data Analysis 3.1: Finding Measures of Position</i>	74
The Mean	75
<i>One Step at a Time: Finding the Mean</i>	76
Three Characteristics of the Mean	77
<i>Statistics in Everyday Life: Means, Medians, and Baseball Salaries</i>	80
<i>Using SPSS: Measures of Central Tendency</i>	81
Choosing a Measure of Central Tendency	84
<i>Social Research and Data Analysis 3.2: The Mean, Mode, and Median</i>	86
Interpreting Statistics: Choosing a Measure of Central Tendency to Describe Income	86
SUMMARY / 88 • SUMMARY OF FORMULAS / 88 • GLOSSARY / 88	
• PROBLEMS / 88	
YOU ARE THE RESEARCHER: THE TYPICAL AMERICAN	94

### Chapter 4 Measures of Variation

97

The Range ( $R$ ) and Interquartile Range ( $Q$ )	99
Computing the Range and Interquartile Range	100
<i>One Step at a Time: Finding the Interquartile Range (<math>Q</math>)</i>	102
<i>Statistics in Everyday Life: Increasing Racial Diversity</i>	102
The Standard Deviation and Variance	103
<i>One Step at a Time: Finding the Standard Deviation (<math>s</math>) and the Variance (<math>s^2</math>)</i>	105
<i>Social Research and Data Analysis 4.1: The Standard Deviation</i>	105



Computing the Standard Deviation: An Additional Example	106
Visualizing Variation: Boxplots	107
<i>Statistics in Everyday Life: Life Expectancy Around the Globe</i>	110
Interpreting the Standard Deviation	111
<i>Social Research and Data Analysis 4.2: Describing Variation</i>	112
<i>Using SPSS: Measures of Variation</i>	113
Interpreting Statistics: The Changing Distribution of Income in the United States	114
<i>Reading Statistics: Pornography and Religiosity</i>	117
SUMMARY / 118 • SUMMARY OF FORMULAS / 118 • GLOSSARY / 119	
• PROBLEMS / 119	
YOU ARE THE RESEARCHER	125

## **Chapter 5 The Normal Curve** **129**

Properties of the Normal Curve	129
<i>Statistics in Everyday Life: Tests of Intelligence</i>	132
Using the Normal Curve	133
<i>One Step at a Time: Finding Z Scores</i>	133
The Normal Curve Table (Appendix A)	134
Finding the Total Area Above and Below a Score	136
<i>One Step at a Time: Finding Areas Above and Below Positive     and Negative Z Scores</i>	138
Finding Areas Between Two Scores	138
<i>One Step at a Time: Finding Areas Between Z Scores</i>	140
<i>Social Research and Data Analysis 5.1 Finding Z Scores and Area Below</i>	140
<i>Social Research and Data Analysis 5.2: Finding Z Scores and Area Between</i>	140
Using the Normal Curve to Estimate Probabilities	141
<i>One Step at a Time: Finding Probabilities</i>	144
<i>Social Research and Data Analysis 5.3: Finding Probabilities</i>	144
<i>Statistics in Everyday Life: Probability Theory in Action</i>	145
SUMMARY / 145 • SUMMARY OF FORMULAS / 145 • GLOSSARY / 145	
• PROBLEMS / 146	
YOU ARE THE RESEARCHER	148

## **Part II Inferential Statistics** **153**

### **Chapter 6 Introduction to Inferential Statistics: Sampling and the Sampling Distribution** **154**

Probability Sampling: Basic Concepts and Terms	155
<i>Statistics in Everyday Life: Advertising and Sampling</i>	156
EPSEM Sampling Techniques	156
<i>One Step at a Time: Drawing a Simple Random Sample</i>	157
<i>One Step at a Time: Drawing a Systematic Random Sample</i>	158
<i>One Step at a Time: Drawing a Stratified Random Sample</i>	159
<i>One Step at a Time: Drawing a Cluster Sample</i>	160
<i>Social Research and Data Analysis 6.1: The American Community Survey</i>	160
<i>Statistics in Everyday Life: Sampling and Telephone Soliciting</i>	161

The Sampling Distribution	161
<i>Statistics in Everyday Life: Using the GSS to Measure Changing American Attitudes</i>	166
The Sampling Distribution: An Additional Example Using the General Social Survey (GSS)	166
<i>Statistics in Everyday Life: Using Surveys to Compare American Attitudes with Other Nations</i>	168
<i>Reading Statistics: Who Offers Extra Credit?</i>	168
Symbols and Terminology	169
SUMMARY / 170 • GLOSSARY / 170 • PROBLEMS / 171	

## Chapter 7 Estimation Procedures

173

Estimation Selection Criteria: Bias and Efficiency	174
Interval Estimation Procedures	177
Interval Estimation Procedures Using Sample Means (Large Samples)	179
<i>One Step at a Time: Constructing Confidence Intervals for Sample Means</i>	181
<i>One Step at a Time: Interpreting the Confidence Interval</i>	181
<i>Social Research and Data Analysis 7.1: Estimating a Population Mean</i>	181
Interval Estimation Procedures Using Sample Proportions (Large Samples)	182
<i>One Step at a Time: Constructing Confidence Intervals for Sample Proportions</i>	183
<i>One Step at a Time: Interpreting the Confidence Interval</i>	184
<i>Social Research and Data Analysis 7.2: Estimating Population Proportions</i>	184
<i>Statistics in Everyday Life: The National Mood</i>	185
Computing Confidence Intervals: A Summary	185
Controlling the Width of Interval Estimates	185
<i>Using SPSS: Constructing Confidence Intervals</i>	187
Interpreting Statistics: Predicting Election Outcomes and Judging the Performance of the President	189
<i>Statistics in Everyday Life: The Classic Election Prediction Error</i>	192
<i>Statistics in Everyday Life: The Big Miss by Pollsters in the 2016 Presidential Election</i>	192
<i>Reading Statistics: Public-Opinion Polls</i>	192
SUMMARY / 194 • SUMMARY OF FORMULAS / 194 • GLOSSARY / 194 • PROBLEMS / 194	
YOU ARE THE RESEARCHER: ESTIMATING THE CHARACTERISTICS OF THE TYPICAL AMERICAN	197

## Chapter 8 Hypothesis Testing I: The One-Sample Case

199

An Overview of Hypothesis Testing	200
<i>Statistics in Everyday Life: Testing Drugs</i>	204
The Five-Step Model for Hypothesis Testing	205
<i>One Step at a Time: Completing Step 4 of the Five-Step Model: Computing Z(obtained)</i>	207
<i>One Step at a Time: Completing Step 5 of the Five-Step Model: Making a Decision and Interpreting Results</i>	208
<i>Statistics in Everyday Life: Hypothesis Testing and Gambling</i>	208
Choosing a One-Tailed or Two-Tailed Test	208

Selecting an Alpha Level	213
<i>Statistics in Everyday Life: Alpha Levels</i>	215
The Student's <i>t</i> Distribution	215
<i>One Step at a Time: Completing Step 4 of the Five-Step Model:</i>	
<i>Computing <math>t(\text{obtained})</math></i>	219
<i>One Step at a Time: Completing Step 5 of the Five-Step Model:</i>	
<i>Making a Decision and Interpreting Results</i>	220
<i>Social Research and Data Analysis 8.1: Testing a Sample Mean for Significance</i>	220
Tests of Hypothesis for Single-Sample Proportions (Large Samples)	221
<i>One Step at a Time: Completing Step 4 of the Five-Step Model:</i>	
<i>Computing <math>Z(\text{obtained})</math></i>	223
<i>One Step at a Time: Completing Step 5 of the Five-Step Model:</i>	
<i>Making a Decision and Interpreting Results</i>	223
<i>Social Research and Data Analysis 8.2: Testing a Sample Proportion</i>	
<i>for Significance</i>	224
SUMMARY / 224 • SUMMARY OF FORMULAS / 225 • GLOSSARY / 225	
• PROBLEMS / 226	

## Chapter 9 Hypothesis Testing II: The Two-Sample Case

229

The One-Sample Case Versus the Two-Sample Case	230
Hypothesis Testing with Sample Means (Large Samples)	231
<i>One Step at a Time: Completing Step 4 of the Five-Step Model:</i>	
<i>Computing <math>Z(\text{obtained})</math></i>	234
<i>One Step at a Time: Completing Step 5 of the Five-Step Model:</i>	
<i>Making a Decision and Interpreting Results</i>	234
<i>Social Research and Data Analysis 9.1: A Test of Significance</i>	
<i>for Sample Means</i>	235
<i>Statistics in Everyday Life: Reading and Math</i>	235
Hypothesis Testing with Sample Means (Small Samples)	236
<i>One Step at a Time: Completing Step 4 of the Five-Step Model:</i>	
<i>Computing <math>t(\text{obtained})</math></i>	238
<i>One Step at a Time: Completing Step 5 of the Five-Step Model:</i>	
<i>Making a Decision and Interpreting Results</i>	238
<i>Using SPSS: Testing Sample Means for Significant Differences</i>	239
Hypothesis Testing with Sample Proportions (Large Samples)	241
<i>One Step at a Time: Completing Step 4 of the Five-Step Model:</i>	
<i>Computing <math>Z(\text{obtained})</math></i>	243
<i>One Step at a Time: Completing Step 5 of the Five-Step Model:</i>	
<i>Making a Decision and Interpreting Results</i>	244
<i>Social Research and Data Analysis 9.2: Testing the Significance of the Difference</i>	
<i>Between Sample Proportions</i>	244
<i>Statistics in Everyday Life: Abortion and Polarization in American Politics</i>	245
The Limitations of Hypothesis Testing: Significance Versus Importance	246
Interpreting Statistics: Is There a Statistically Significant Gender Gap in Income?	247
<i>Statistics in Everyday Life: Is the Gender Gap Decreasing?</i>	249
<i>Reading Statistics: Comparing Ethnic and Racial Social Distance by Gender</i>	250

SUMMARY / 250 • SUMMARY OF FORMULAS / 250 • GLOSSARY / 251
• PROBLEMS / 251
YOU ARE THE RESEARCHER: GENDER GAPS AND SUPPORT FOR
TRADITIONAL GENDER ROLES 256

## **Chapter 10 Hypothesis Testing III: The Analysis of Variance 260**

The Logic of the Analysis of Variance 261
The Computation of ANOVA 262
An Example of Computing the Analysis of Variance 264
A Test of Significance for ANOVA 266
<i>One Step at a Time: Completing Step 4 of the Five-Step Model:</i>
<i>Computing F(obtained) 268</i>
<i>One Step at a Time: Completing Step 5 of the Five-Step Model:</i>
<i>Making a Decision and Interpreting Results 269</i>
<i>Statistics in Everyday Life: What Causes Obesity? 269</i>
<i>Social Research and Data Analysis 10.1: The Analysis of Variance 269</i>
<i>Using SPSS: The ANOVA Test 271</i>
An Additional Example of Computing and Testing the Analysis of Variance 272
The Limitations of the Test 275
Interpreting Statistics: Does Sexual Activity Vary by Age? 276
<i>Reading Statistics: Religion and Sex Education 279</i>
SUMMARY / 279 • SUMMARY OF FORMULAS / 280 • GLOSSARY / 280
• PROBLEMS / 281
YOU ARE THE RESEARCHER: WHY ARE SOME PEOPLE LIBERAL
(OR CONSERVATIVE)? 285

## **Chapter 11 Hypothesis Testing IV: Chi Square 289**

Bivariate Tables 290
The Logic of Chi Square 291
The Computation of Chi Square 292
<i>One Step at a Time: Computing Expected Frequencies (<math>f_e</math>) 295</i>
<i>One Step at a Time: Computing Chi Square 295</i>
The Chi Square Test for Independence 295
Column Percentages 297
<i>Social Research and Data Analysis 11.1: The Chi Square Test 298</i>
<i>One Step at a Time: Computing Column Percentages 299</i>
<i>Statistics in Everyday Life: The Gender Gap 299</i>
The Chi Square Test: An Additional Example 300
<i>Using SPSS: Crosstabs and Chi Square 302</i>
Chi Square and the Goodness-of-Fit Test 303
The Limitations of the Chi Square Test 305
<i>Statistics in Everyday Life: Who Uses Social Networking Sites? 307</i>
Interpreting Statistics: Family Values and Social Class 307
<i>Social Research and Data Analysis 11.2: Racism and Free Speech 309</i>
<i>Reading Statistics: Awareness of Support for LGBT Communities Among Youth 310</i>

SUMMARY / 310 • SUMMARY OF FORMULAS / 311 • GLOSSARY / 311
• PROBLEMS / 312
YOU ARE THE RESEARCHER: UNDERSTANDING POLITICAL BELIEFS 315

## **Part III Bivariate Measures of Association 321**

### **Chapter 12 Bivariate Association for Nominal- and Ordinal-Level Variables 322**

Association and Bivariate Tables 323
Three Characteristics of Bivariate Associations 324
<i>Social Research and Data Analysis 12.1: The Importance of Percentages and Avoiding Errors of Interpretation</i> 329
<i>Statistics in Everyday Life: Discrimination in Football?</i> 331
Measures of Association: Introduction 331
Measures of Association for Nominal Variables 331
<i>One Step at a Time: Computing Cramer's V</i> 332
<i>One Step at a Time: Interpreting Cramer's V</i> 332
<i>One Step at a Time: Computing Lambda</i> 336
<i>One Step at a Time: Interpreting Lambda</i> 336
<i>Social Research and Data Analysis 12.2: Using Cramer's V and Lambda</i> 337
<i>Statistics in Everyday Life: Confidence in Higher Education</i> 337
Measures of Association for Collapsed Ordinal Variables 338
<i>One Step at a Time: Computing Gamma</i> 341
<i>One Step at a Time: Interpreting Gamma</i> 341
<i>Social Research and Data Analysis 12.3: Interpreting Direction with Ordinal-Level Variables</i> 342
<i>Using SPSS: Crosstabs and Gamma</i> 342
A Measure of Association for Continuous Ordinal Variables: Spearman's Rho ( $r_s$ ) 343
<i>One Step at a Time: Computing Spearman's Rho</i> 346
<i>One Step at a Time: Interpreting Spearman's Rho</i> 346
Interpreting Statistics Age and Concern for the Environment 347
<i>Reading Statistics: Bivariate Tables</i> 348
SUMMARY / 348 • SUMMARY OF FORMULAS / 348 • GLOSSARY / 349
• PROBLEMS / 349
YOU ARE THE RESEARCHER: UNDERSTANDING
POLITICAL BELIEFS, PART II 353

### **Chapter 13 Association Between Variables Measured at the Interval-Ratio Level 357**

Scatterplots and Regression Analysis 358
<i>Statistics in Everyday Life: Who Does the Housework?</i> 361
<i>Using SPSS: Scatterplots</i> 362
The Regression Line 364
Computing and Interpreting the Regression Coefficients ( $a$ and $b$ ) 365
<i>One Step at a Time: Computing the Slope (<math>b</math>)</i> 367
<i>One Step at a Time: Computing the Y Intercept (<math>a</math>)</i> 368

Predicting Scores on $Y$ with the Least-Squares Regression Line	368
<i>One Step at a Time: Using the Regression Line to Predict Scores on <math>Y</math></i>	368
The Correlation Coefficient (Pearson's $r$ )	369
<i>One Step at a Time: Computing Pearson's <math>r</math></i>	370
Interpreting the Correlation Coefficient $r^2$	370
<i>One Step at a Time: Interpreting the Strength and Direction of Pearson's <math>r</math></i>	373
<i>Social Research and Data Analysis 13.1: Computing the Regression Coefficients and Pearson's <math>r</math></i>	374
The Correlation Matrix	375
<i>Using SPSS: Producing a Correlation Matrix</i>	375
<i>Reading Statistics: Predicting Knowledge Sharing in Organizations, Part 1</i>	377
Correlation, Regression, Level of Measurement, and Dummy Variables	377
<i>Statistics in Everyday Life: Teen Pregnancies</i>	378
<i>Using SPSS: Conducting a Regression Analysis with a Dummy Variable</i>	379
Testing Pearson's $r$ for Significance	379
<i>Social Research and Data Analysis 13.2: Fertility and Women's Education</i>	380
<i>Statistics in Everyday Life: Correlation, Causation, and Cancer</i>	382
Interpreting Statistics: The Correlates of Crime	382
SUMMARY / 384 • SUMMARY OF FORMULAS / 385 • GLOSSARY / 385	
• PROBLEMS / 386	
YOU ARE THE RESEARCHER: WHO WATCHES TV?	
WHO SUCCEEDS IN LIFE?	390

## Part IV Multivariate Techniques 397

### Chapter 14 Elaborating Bivariate Tables 398

Controlling for a Third Variable	399
Interpreting Partial Tables	401
<i>Statistics in Everyday Life: Spuriousness</i>	404
<i>Statistics in Everyday Life: Interaction</i>	406
Partial Gamma ( $G_p$ )	407
<i>One Step at a Time: Computing Partial Gamma (<math>G_p</math>)</i>	409
<i>Social Research and Data Analysis 14.1: Elaborating Bivariate Tables</i>	410
Where Do Control Variables Come From?	410
<i>Using SPSS: Crosstabs and Measures of Association</i>	411
The Limitations of Elaborating Bivariate Tables	412
Interpreting Statistics: Another Look at Concern for the Environment	413
<i>Reading Statistics: Predicting Knowledge Sharing Attitudes, Part 2</i>	415
SUMMARY / 415 • SUMMARY OF FORMULAS / 416 • GLOSSARY / 416	
• PROBLEMS / 416	
YOU ARE THE RESEARCHER: A MULTIVARIATE ANALYSIS	
OF ATTITUDES ON ABORTION	421

### Chapter 15 Partial Correlation and Multiple Regression and Correlation 424

Partial Correlation	425
<i>One Step at a Time: Computing the Partial Correlation Coefficient</i>	430
<i>One Step at a Time: Interpreting the Partial Correlation Coefficient</i>	430

<i>Social Research and Data Analysis 15.1: Education and Income</i>	431
<i>Social Research and Data Analysis 15.2: Using SPSS to Determine If the Relationship Between Fertility and Women's Education Is Affected by Affluence</i>	431
Multiple Regression: Predicting the Dependent Variable	431
<i>One Step at a Time: Computing Partial Slopes</i>	434
<i>One Step at a Time: Interpreting Partial Slopes</i>	434
<i>One Step at a Time: Computing the Y Intercept</i>	435
<i>One Step at a Time: Using the Multiple Regression Line to Predict Scores on Y</i>	435
Multiple Regression: Assessing the Effects of the Independent Variables	436
<i>One Step at a Time: Computing Beta-Weights (<math>\beta</math>)</i>	437
<i>One Step at a Time: Interpreting Beta-Weights (<math>\beta</math>)</i>	437
Multiple Correlation	438
<i>One Step at a Time: Computing the Coefficient of Multiple Determination (<math>R^2</math>)</i>	440
<i>One Step at a Time: Interpreting (<math>R^2</math>)</i>	440
<i>Using SPSS: Regression Analysis</i>	440
<i>Social Research and Data Analysis 15.3: Multiple Regression and Correlation</i>	441
Interpreting Statistics: Another Look at the Correlates of Crime	443
<i>Statistics in Everyday Life: Statistics and Baseball</i>	446
The Limitations of Multiple Regression and Correlation	446
<i>Reading Statistics: Do Sociology Courses Increase Empathy in Students?</i>	447
SUMMARY / 449 • SUMMARY OF FORMULAS / 449 • GLOSSARY / 449	
• PROBLEMS / 450	
YOU ARE THE RESEARCHER: A MULTIVARIATE ANALYSIS OF TV VIEWING AND SUCCESS	454
 <i>Appendix A Area Under the Normal Curve</i>	 461
<i>Appendix B Distribution of t</i>	465
<i>Appendix C Distribution of Chi Square</i>	467
<i>Appendix D Distribution of F</i>	469
<i>Appendix E Using Statistics: Ideas for Research Projects</i>	471
<i>Appendix F An Introduction to SPSS</i>	477
<i>Appendix G Codebooks for All Data Sets</i>	483
 <i>Answers to Odd-Numbered End-of-Chapter Problems and Cumulative Exercises</i>	 497
 <i>Glossary</i>	 513
 <i>Index</i>	 519





# Preface

Society has become increasingly data driven. At colleges and universities, statistics are an essential part of sociology and other social sciences (including political science, social work, psychology, public administration, criminal justice, urban studies, and others). In social policy work, researchers are challenged to use data (on poverty, health care, policing, and other topics) to demonstrate why policy reform will be beneficial to society. In the private sector, and in sports, decisions are increasingly based on data analytics.

This means that understanding data has become an essential skill that college students can bring to the job market when they graduate. Nevertheless, learning statistics is challenging, especially for students that do not have an extensive background in mathematics. Social science students, in particular, possess a wide range of skills and interests but still find themselves puzzled about the relevance of statistics for them, and frequently, this leads to math anxiety. This text introduces statistical analysis for the social sciences in a way that is intended to reduce those anxieties.

## The Authors and the New Title of the Book

We begin by telling you a little bit about the authors and the new approach they have taken to this edition. **Joseph Healey** taught sociology and statistics for more than 40 years. His research and teaching interests centered on ethnic and race relations, immigration, sports, violence, and globalization. He began this project because of his dissatisfaction with texts that were either too mathematical and technical or too simplistic. He sought to find a middle approach that was accessible to beginning students and that prepared them to appreciate the professional literature and to do actual social research themselves.

New to the 11th edition is the addition of **Christopher Donoghue** as coauthor. Chris is a sociologist with research interests in peer aggression and bullying, school climate, ethnic and racial social distance, and sexual behavior among adolescents. He has also conducted extensive research on the quality of care in nursing homes. As a student, he used a much earlier edition of this book and since he began teaching the course in 2001, it is the only book he has used. Although he has taught the course for many years now, every semester is different (students come with new needs and interests, trends in using data change quite often). Changing the course to add new techniques, technologies, and strategies is something that he greatly enjoys. Joe and Chris also share an interest in effectively teaching statistics to students that lack confidence in mathematics, so their collaboration on this text has been a natural fit.

Together, Joe and Chris set out to make the 11th edition of the book more current and reflective of the increased emphasis in society on data and evidence-based decision making. The new title, **Statistics: A Tool for Social Research and Data Analysis**, is indicative of the book's new approach to use real data and demonstrate real applications in all chapters. Although the previous editions of the book used real data, they also contained many boxes on Applied Statistics that used fictional data. In the latest edition, these boxes have been replaced by boxes on **Social Research and Data Analysis**, nearly all of which contain real problems with actual data. Many other updates were made to the text as well, with the consistent goal of infusing both real and current data to the learning experience.

## Who Is This Book For?

Most undergraduate students of sociology and the other social sciences take a course in social statistics. This book is rich with examples from those fields and it is designed for that audience, but its many additional examples from social policy research, market research and sports, and its applied orientation make it flexible enough to be used in statistics courses for all kinds of undergraduate majors as well as graduate courses.

Many students dread that first course in statistics, especially if they have not had success in mathematics in the past. This book was designed with that sort of student in mind. Since 1984, students have been using this text to learn about statistics in an approachable and straightforward fashion. The book contains simple instructions, language that is easy to understand, and data from research and everyday life that students will enjoy reading about. The book makes few assumptions about mathematical background. In fact, if you can *add*, *subtract*, *multiply*, and *divide*, you will find that almost all of the operations in the text can be accomplished without much additional mathematical learning. They just take some concentration and the willingness to learn new concepts and methods.

The book is written at an intermediate level and its goal is to show the relevance and value of statistics for the social sciences. We emphasize interpretation and understanding statistics in the context of social science research, but we 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. The text does not claim that statistics are always “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 frustrate the learning experience at this level.

## Changes to the 11th Edition

Previous users will find that the structure of the book and the instructional content have not changed. What is new is the extensive addition of new data on current problems throughout the text from sociology and the other social sciences (e.g., political science, social work, and gerontology), market research, social

policy, sports, and more. The influx of data analytics to professional industries is also discussed and students are encouraged to learn more by exploring statistical software packages such as SPSS, SAS, and R.

The following are the most important changes to this edition:

- Current examples and information can be found in all chapters on topics such as the 2016 presidential election in the United States, the mass shooting epidemic, contemporary uses of the term “racism,” debates about the appropriateness of abstinence teaching in sexual education, the “fake news” phenomenon, and more.
- Social Research and Data Analysis boxes now contain examples in which students are encouraged to use the formulas in the chapter to explore real and current data.
- New boxes on Reading Statistics have been added, containing interesting findings from real academic and professional studies.
- The Pew Research Center’s 2018 Core Trends data set has been added to provide students with more variables to explore than in previous editions.
- All of the data on current events, the SPSS exercises, and the Statistics in Everyday Life boxes have been updated.
- New end-of-chapter problems have been added on the topics of Market Research and Social Policy.
- In Chapter 15, new terms for multiple regression statistics have been added (e.g., unstandardized regression coefficients and constants) to make the chapter more consistent with academic research articles and more usable to students.
- New commentary now appears in Chapter 3 on bimodal distributions, Chapter 4 on skewness and kurtosis, and Chapter 6 on sampling frames.
- Several new Step-by-Step boxes have been added to Chapter 6 to teach students how to employ different random sampling techniques.
- The Using Statistics boxes now direct students to specific pages in the chapters where the subject material is used to solve an interesting problem or to address an important topic.
- SPSS problems have been added to the end-of-chapter problems throughout the text. In some chapters (e.g., Chapters 13 and 15), the SPSS problems replace problems using hand calculators.
- The data sets used in the text have been expanded and updated. The data sets 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 data sets are available for downloading at the Web site for this text: **www.cengagebrain.com**; they include:
  - A General Social Survey (GSS) data set (*GSS2018.sav*), which has been updated to 2018.
  - A smaller version of the 2018 GSS (*GSS2018 Sample*) for students using the student version of SPSS. This data set has only 50 variables and fewer than 1500 cases.

- The Pew Research Center’s 2018 Core Trends Data Set.
- A data set that includes census and crime data on the 50 states (*States.sav*).
- A data set that includes demographic data for 99 nations (*International.sav*).
- A fourth data set (*CrimeTrends84-17.sav*) is used only for the graphing exercises in Chapter 2.

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

## Goal of the Text

Although the book is evolving, its goal is still to introduce students to basic statistics and demonstrate their usefulness in social research of all kinds. The statistically literate person understands the role of statistics in the research process, is competent to perform basic calculations, and can appreciate any analyses of data that she or he may encounter in everyday life. Let’s consider these three competencies in a little more detail.

**1. An Appreciation of Statistics.** Appreciating statistics means understanding their relevance and being capable of employing statistical tools when they are needed, and in the right situations. This textbook develops these qualities in several ways.

- *The relevance of statistics.* Throughout the text, each example 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 “Social Research and Data Analysis” boxes, each of which illustrates some specific way statistics can be used to answer questions, such as those posed in the “Using Statistics” feature that opens every chapter.
- *Employing techniques and strategies.* At the end of each chapter, students are challenged to use what they have learned to solve problems relevant to specific fields: **SOC for sociology, SW for social work, PS for political science, SP for social policy (new to this edition), MR for market research (new to this edition), CJ for criminal justice, PA for public administration, and GER for gerontology.** These problems show the relevance of statistics to students with a wide array of academic and career interests.
- *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, we have been careful to express the meaning of the statistics in terms of the original research question. To provide experience, the Using SPSS boxes and the end-of-chapter problems call for an interpretation of the statistics that are calculated. Students can also learn these lessons from the “Interpreting Statistics” sections that appear in many chapters.

- *Using statistics: Ideas for research projects.* Appendix E and the You Are the Researcher boxes offer ideas for independent data-analysis projects for students. The projects require students to use SPSS to analyze a data set. 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 to exercise their ability to 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 we 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.
    - *Cumulative exercises* are included at the end of each Part of the book to provide practice in choosing, computing, and analyzing statistics. These exercises present only data sets and research questions. Students must choose appropriate statistics as part of the exercise.
    - *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.** Comprehending and critically appraising statistical reports is another essential skill that can come out of a basic statistics course. The development of this skill is a particular problem for many students since (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. To help bridge this gap, we have included a series of boxes labeled “**Reading Statistics.**” In each such box, we briefly describe the reporting style typically used for the statistic in question and try to alert students about what to expect when they approach the professional literature. These inserts have been updated in this edition and most include excerpts from the research literature that illustrate how statistics are actually applied and interpreted by social scientists.

## 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. 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 front cover and a glossary of symbols inside the back cover 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, we 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.
- *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 data sets 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 the 2018 General Social Survey, the Pew Research Center’s 2018 Core Trends data set, a data set that includes census and crime data for the 50 states, and a data set that includes demographic data for 99 nations. These data sets will give students the opportunity to practice their statistical skills on “real-life” data. All databases are described in Appendix G.

# Supplements

Several ancillaries accompany the text:

**MindTap™: The Personal Learning Experience** The redesigned MindTap for *Statistics: A Tool for Social Research and Data Analysis*, 11th Edition, from Cengage represents a new approach to a highly personalized, online learning platform. MindTap combines all of a student's learning tools, readings, and multimedia activities into a "Learning Path" that guides students through their statistics course. Four new, highly interactive activities challenge students to think critically by exploring, analyzing, and creating content, developing their sociological imagination through personal, local, and global lenses.

*MindTap for Statistics: A Tool for Social Research and Data Analysis* is easy to use and saves instructors time by allowing you to

- Break course content down into manageable modules to promote personalization, encourage interactivity, and ensure student engagement.
- Customize the course—from tools to text—and make adjustments "on the fly," making it possible to intertwine breaking news into your lessons and incorporate today's teachable moments.
- Bring interactivity into learning through the integration of multimedia assets (apps from Cengage and other providers) and numerous in-context exercises and supplements; student engagement will increase, leading to better student outcomes.
- Track students' use, activities, and comprehension in real time, which provides opportunities for early intervention to influence progress and outcomes. Grades are visible and archived so students and instructors always have access to current standings in the class.
- Assess knowledge throughout each section: after readings, and in automatically graded activities and assignments.
- A new digital implementation guide will help you integrate the new MindTap Learning Path into your course.
- Learn more at [www.cengage.com/mindtap](http://www.cengage.com/mindtap).
- *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 <http://login.cengage.com>.
- *Cengage Learning Testing Powered by Cognero®*. Cengage Learning Testing powered by Cognero is a flexible online system that allows instructors to author, edit, and manage Test Bank content and quickly create multiple test versions. You can deliver tests from your LMS, your classroom—or wherever you want.



## Acknowledgments

*Joseph F. Healey*

I would like to thank the instructors who made statistics understandable to me (Professors Satoshi Ito, Noelle 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.

I want to thank Professor Christopher Donoghue for his countless contributions as coauthor of the 11th edition. His understanding of the material and his years of experience in teaching and research have greatly strengthened this text, and I look forward to many years of collaboration on this project.

I also 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), my wife Patricia A. Healey, the next generation (my sons, Kevin and Christopher, my daughters-in-law, Jennifer and Jessica, my stepson Christopher Schroen, and stepdaughters Kate Cowell and Jennifer Schroen), and the youngest generation (Benjamin, Caroline, and Isabelle Healey and Abigail Cowell).

*Christopher Donoghue*

I owe a great deal of debt to three mentors that I have been fortunate to have in Social Statistics. The first is Dr. Rosemary Cooney of Fordham University. I like to say that Rosemary taught me the “language of statistics.” To this day, I can hear her voice whenever I am teaching the course. Dr. Joseph Healey was also a mentor to me in statistics even though I did not meet him in person until 2013. He wrote this book to help students who struggled with mathematics, and that is exactly the kind of student I was as an undergraduate in the 1990s. This book made statistics easy for me and it helped me immensely in communicating the subject matter to my students. Finally, Dr. Yong Wang of Montclair State University has been a constant and reliable source of information about statistics for me over the last several years. Yong has pushed me to incorporate many new techniques that have improved my courses. He is also the first person I talk to when I need the answer to a question about statistics.

Joe and I also would like to express a special thanks to our Writer's Assistant, Taulant Asani. Taulant was a great reviewer of our work and he was excellent at checking our math in Excel. We are also in debt to Stephen Shahin, and undergraduate student Montclair State University. Stephen was a careful reader of the prior edition. He caught a number of our errors and he made great suggestions for the current edition.

I would also like to thank my wife, Lisa Donoghue, my daughter, Amanda Donoghue, and my mother Teresa Donoghue. All three of them have supported me throughout my academic career, and in this endeavor, they have been a constant source of motivation that I could not do without.

Finally, we are 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 Appendixes B, C, and D from their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974).



# Prologue: Basic Mathematics Review

You will probably be relieved to hear that this text does not place a great emphasis on mathematical computations. While you will encounter many numbers to work with and formulas to solve, the main focus 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 sample problems have been simplified so that the computations will not impede your understanding of the statistics themselves.

Statistics do involve numbers, however, so you cannot avoid the mathematics altogether. If you have anxieties about math, we recommend you focus on the four simplest operations with which you are already familiar: *adding*, *subtracting*, *multiplying*, and *dividing*. When it comes down to it, every formula in this book can be completed by carrying out those four simple tasks. The greater challenges in the work of statistics are instead about figuring out which of these tasks to employ, and when to employ them. So to make your work easier, we have included boxes called “Step by Step” that tell you exactly which tasks to perform and when to perform them. We recommend you try these boxes out whenever a formula gives you trouble.

## 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, if you have experience using Microsoft Excel or Google Sheets, you will find that these software packages can help you even more because they will enable you to perform simple tasks (like adding and multiplying) over and over again in a clear and efficient manner. In the text, you will also learn how to use SPSS for this purpose. One benefit of SPSS is that it enables you to perform more sophisticated techniques with a simple point and click method, but unfortunately, it does not show you how the statistics are computed. For this reason, there is also a benefit to using Microsoft Excel or Google Sheets because these packages require you to set the math up yourself.

## 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 (e.g., income) or relatively few (e.g., gender). We 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 a 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. We 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:

$$\begin{aligned} a \times b \\ a \cdot b \\ a * b \\ ab \\ a(b) \\ (a)(b) \end{aligned}$$

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  $\Sigma X_i$ .

“ $\Sigma$ ” is uppercase Greek letter sigma and stands for “the summation of.” So the combination of symbols  $\Sigma X_i$  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  $\Sigma$  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  $\Sigma X^2$ , 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:

Symbol	Operations
$\Sigma X_i$	Add the scores
$\Sigma X_i^2$	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 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 we (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. Our practice is to work in as much accuracy as our 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 end. If a set of calculations is lengthy and requires the reporting of intermediate sums or subtotals, we will round the subtotals off to two places as we 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:

$$\sum X_i =$$

$$\sum X_i^2 =$$

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

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.  $\frac{(4 - 3) + (7 + 2)(3)}{(4 + 5)(10)}$

d.  $\frac{22 + 44}{15/3}$

## 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

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

The second expression is the “sum of the squared scores.” So

$$\sum X_i^2 = (50)^2 + (55)^2 + (60)^2 + (65)^2 + (70)^2$$

$$\sum X_i^2 = 2500 + 3025 + 3600 + 4225 + 4900$$

$$\sum X_i^2 = 18,250$$

The third expression is “the sum of the scores, squared”:

$$(\sum X_i)^2 = (50 + 55 + 60 + 65 + 70)^2$$

$$(\sum X_i)^2 = (300)^2$$

$$(\sum X_i)^2 = 90,000$$

Remember that  $\sum X_i^2$  and  $(\sum X_i)^2$  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



# 1

## Introduction

### LEARNING OBJECTIVES

By the end of this chapter, you will be able to:

1. Identify the key stages and terms in social scientific research.
2. Distinguish between descriptive and inferential statistics.
3. Provide examples of discrete and continuous variables.
4. Describe three levels of measurement and cite examples of each.

### USING STATISTICS

One of the most important themes of this text is that statistics are useful tools for analyzing and understanding information and for communicating our conclusions to others.

To stress this theme, each chapter will begin with a list of situations in which statistics can be applied to good advantage. In some chapters, you can find the page numbers on which we use statistics for these purposes.

Statistics can be used to

- Describe the relationship between Internet use and social class. **See page 8**
- Make estimates about the percentage of Americans that are in favor of same-sex marriage and those that are against it. **See page 9**
- Measure the percentage of Americans that identify as members of the lower class over time. **See page 13**
- Compare the cost of living (housing prices and rents, the cost of groceries and gas, health care, and so forth) between different localities (cities, states, and even nations).

## Why Study Statistics?

Students sometimes wonder about the value of studying statistics. What, after all, do numbers and formulas have to do with understanding people and society? The value of statistics will become clear as we move from chapter to chapter, but, for now, we can demonstrate the importance of statistics by considering **research** or disciplined and careful inquiries. Scientists conduct

research to answer questions, examine ideas, and test theories. Research can take numerous forms, and statistics are relevant for **quantitative research** projects, or projects that collect information in the form of numbers or **data**. **Statistics** are mathematical techniques used by social scientists to analyze data in order to answer questions and test theories.

What is so important about learning how to manipulate data? On one hand, some of the most important and enlightening works in the social sciences do not utilize statistics at all. There is nothing magical about data and statistics. The mere presence of numbers guarantees nothing about the quality of a scientific inquiry.

On the other hand, data can be the most trustworthy information available to the researcher and, consequently, deserve special attention. Data that have been carefully collected and thoughtfully analyzed can be the strongest, most objective foundations for building theory and enhancing our understanding of the social world.

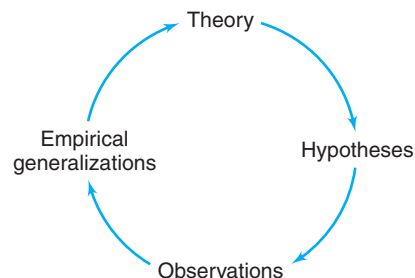
Let us be clear about one point: It is never enough merely to gather data (or, for that matter, any kind of information). Even the most carefully collected numerical information does not and cannot speak for itself. The researcher must be able to use statistics effectively to organize, evaluate, and analyze the data. Without a good understanding of statistical analysis, the researcher will be unable to make sense of the data. Without an appropriate application of statistics, data are useless.

Statistics are an indispensable tool for the social sciences. They provide the scientist with some of the most useful techniques for evaluating ideas, testing theory, and making evidence-based decisions. The next section describes the relationships among theory, research, and statistics in more detail.

## The Role of Statistics in Scientific Inquiry

Figure 1.1 graphically represents the role of statistics in the research process. The diagram is based on the thinking of Walter Wallace and illustrates a useful conception of how any scientific enterprise grows and develops. One point the diagram makes is that scientific theory and research continually shape each other. Statistics are one of the most important means by which research and theory interact. Let's take a closer look at the process.

**FIGURE 1.1 The Wheel of Science**



Source: Adapted from Walter Wallace, *The Logic of Science in Sociology* (Chicago: Aldine-Atherton, 1971).

## A JOURNEY THROUGH THE SCIENTIFIC PROCESS

Because Figure 1.1 is circular, with no beginning or end, we could start our journey at any point. For the sake of convenience, let's begin at the top, with theory, and follow the arrows around the circle. A **theory** is an explanation of the relationships among phenomena. People naturally (and endlessly) wonder about problems in society (such as prejudice, poverty, child abuse, and serial murders), and they develop explanations (“low levels of education cause prejudice”) in their attempts to understand. Unlike our everyday, informal explanations, scientific theory is subject to a rigorous testing process. Let's take prejudice as an example to illustrate how the research process works.

**Theory.** Why are some people prejudiced against other groups? One possible answer to this question is provided by a theory of prejudice stated more than 70 years ago by social psychologist Gordon Allport and tested on a number of occasions since that time.<sup>1</sup>

The theory states that prejudice will decrease in situations in which members of different groups have equal status and cooperate with each other to work toward mutual goals. The more equal and cooperative the contact, the more likely people will see each other as individuals and not as representatives of a particular group. For example, we might predict that members of a racially mixed athletic team who work together to achieve victory will experience a decline in prejudice. However, when different groups compete for jobs, housing, or other resources, prejudice will tend to increase. Allport's theory is not a complete explanation of prejudice, of course, but it will serve to illustrate a sociological theory.

**Variables and Causation.** Note that Allport's theory is stated in terms of causal relationships between two variables. A **variable** is any trait that can change values from case to case; examples include gender, age, income, and political party affiliation. A theory may identify some variables as causes and others as effects or results. In the language of science, the causes are **independent variables** and the effects or results are **dependent variables**. In our theory, equal-status contact would be the independent variable (or the cause) and level of prejudice would be the dependent variable (the result or effect). In other words, the theory argues that the level of prejudice *depends on* (or is caused by) the extent to which a person participates in equal-status, cooperative contacts with members of other groups.

---

<sup>1</sup>Allport, Gordon, 1954. *The Nature of Prejudice*. Reading, MA: Addison-Wesley. This theory is often called “the contact hypothesis.” For recent attempts to test this theory, see: McLaren, Lauren. 2003. “Anti-Immigrant Prejudice in Europe: Contact, Threat Perception, and Preferences for the Exclusion of Migrants.” *Social Forces*, 81:909–937; Pettigrew, Thomas. 1997. “Generalized Intergroup Contact Effects on Prejudice.” *Personality and Social Psychology Bulletin*, 23:173–185; and Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. 2011. “Recent Advances in Intergroup Contact Theory.” *International Journal of Intercultural Relations*, 35: 271–280.

Diagrams can be a useful way of representing the relationships between variables:

$$\begin{aligned} &\text{Equal-status contact} \rightarrow \text{Prejudice} \\ &\text{Independent variable} \rightarrow \text{Dependent variable} \\ &X \rightarrow Y \end{aligned}$$

The arrow represents the direction of the causal relationship, and  $X$  and  $Y$  are general symbols for the independent and dependent variables, respectively.

**Hypotheses.** So far, we have a theory of prejudice and independent and dependent variables. Is the theory true or false? To find out, we need to compare our theory with the facts; we need to do some research. Our next steps would be to define our terms and ideas. One problem we often face in research is that theories are complex and abstract, and we need to be very specific to conduct a valid test. Often, we do this by deriving a hypothesis from the theory: A **hypothesis** is a statement about the relationship between variables that is specific and exact.

For example, if we wished to test Allport's theory, we would have to say exactly what we mean by prejudice and we would need to describe "equal-status, cooperative contact" in detail. We would also review the research literature to help develop and clarify our definitions and our understanding of these concepts.

As our hypothesis takes shape, we begin the next step of the research process, during which we decide exactly how we will gather data. We must decide how cases will be selected and tested, how variables will be measured, and a host of related matters. Ultimately, these plans will lead to the observation phase (the bottom of the wheel of science), where we actually measure social reality. Before we can do this, we must have a clear idea of what we are looking for and a well-defined strategy for conducting the search.

**Making Observations and Using Statistics.** To test Allport's theory of prejudice, we might begin with people from different racial or ethnic groups. We might place some subjects in cooperative situations and others in competitive situations. We would need to measure levels of prejudice before and after each type of contact. We might do this by administering a survey that asks subjects to agree or disagree with statements such as "Greater efforts must be made to racially integrate the public school system." Our goal would be to see whether the people exposed to the cooperative-contact situations actually become less prejudiced.

Now, finally, we come to the use of statistics in our research. During the observation phase, we will collect a great deal of numerical information or data. If we have a sample of 100 people, we will have 200 completed surveys measuring prejudice: 100 completed before the contact situation and 100 filled out afterward. Try to imagine dealing with 200 completed surveys. If we ask each respondent just five questions to measure their prejudice, we will have a total of 1000 separate pieces of information to deal with. What will we do? We'll have to organize and analyze this information, and statistics will become very helpful at this point. They will supply us with many ideas about "what to do" with the data. We will begin to look at some of the options in the next chapter; for now, let us stress two points about statistics.

First, statistics are crucial. Simply put, without statistics, quantitative research is impossible. We need statistics to analyze data and to shape and refine our theories of the social world.

Second, and somewhat paradoxically, the role of statistics is limited. As Figure 1.1 makes clear, scientific research proceeds through several mutually interdependent stages. Statistics become directly relevant only at the end of the observation stage. Before any statistical analysis can be legitimately applied, however, earlier phases must have been successfully completed. If the researcher has asked poorly conceived questions or has made serious errors of design or method, then even the most sophisticated statistical analysis is valueless. As useful as they can be, statistics cannot substitute for rigorous conceptualization, detailed and careful planning, or creative use of theory. Statistics cannot salvage a poorly conceived or designed research project. They cannot make sense out of garbage.

On the other hand, inappropriate statistical applications can limit the usefulness of an otherwise carefully done project. Only by successfully completing *all* phases of the process can a quantitative research project hope to contribute to understanding. A reasonable knowledge of the uses and limitations of statistics is as essential to the education of the social scientist as is training in theory and methodology.

**Empirical Generalizations.** Our statistical analysis would focus on assessing our theory, but we would also look for other trends in the data. For example, if we found that equal-status, cooperative contact reduces prejudice in general, we might go on to ask whether the pattern applies to males as well as to females and to the well educated as well as to the poorly educated. As we probed the data, we might begin to develop some generalizations based on the empirical patterns we observe. For example, what if we found that contact reduced prejudice for younger respondents but not for older respondents? Could it be that younger people are less “set in their ways” and have attitudes and feelings that are more open to change? As we developed tentative explanations, we would begin to revise or elaborate the theory that guides the research project.

## STATISTICS

### In Everyday Life

#### Push Polls

Political campaigns sometimes use “push polls” to sway public opinion. These polls are designed to influence opinions, sometimes by starting or circulating rumors and innuendo. For example, they may try to brand an opponent as untrustworthy by asking questions like “Would you still support the candidate if you found out that he is an alcoholic?” Even when completely fabricated, such a question can create negative associations in the minds of voters. Statistics may be used to “analyze” data gathered by push polls (or by marketing campaigns that use similar techniques), but the results have little or no scientific validity.

**A New Theory?** If we changed our theory because of our empirical generalizations, a new research project would probably be needed to test the revised theory, and the wheel of science would begin to turn again. We (or perhaps other researchers) would go through the entire process once again with this new—and, hopefully, improved—theory. This second project might result in further revisions that would require still more research, and the wheel of science would continue to turn as long as scientists were able to suggest revisions or develop new insights. Every time the wheel turned, our understanding of the phenomenon under consideration would (hopefully) improve.

**Summary: The Dialog Between Theory and Research.** This description of the research process does not include white-coated, clipboard-carrying scientists who, in a blinding flash of inspiration, discover some fundamental truth about reality and shout “Eureka!” The truth is that, in the normal course of science, we rarely are in a position to declare a theory true or false. Rather, evidence will gradually accumulate over time, and ultimate judgments of truth will be the result of many years of hard work, research, and debate. It is for this reason that social scientists usually report that their results point to “evidence” for something, rather than saying that it “proves” something.

Let’s briefly review our imaginary research project. We began with a theory and examined the various stages of the research project that would be needed to test that theory. We wound up back at the top of the wheel, ready to begin a new project guided by a revised theory. We saw how theory can motivate research and how our observations might cause us to revise theory and, thus, motivate a new research project: Theory stimulates research and research shapes theory. This constant interaction is the lifeblood of science and the key to enhancing our understanding of the social world.

The dialog between theory and research occurs at many levels and in multiple forms. Statistics are one of the most important links between these two realms. Statistics permit us to analyze data, to identify and probe trends and relationships, to develop generalizations, and to revise and improve our theories. As you will see throughout this text, statistics are limited in many ways. They are also an indispensable part of the research enterprise, an essential tool for conducting research and shaping theory. (*For practice in describing the relationship between theory and research and the role of statistics in research, see problems 1.1 and 1.2.*)

## The Goals of This Text

Society is becoming increasingly data driven. Social scientific researchers, as well as other professionals that work in social policy, market research, sports, and many other fields, use data to aide their decision making all the time. The goals of this book are to assist students in becoming skilled at using basic statistics as well as to introduce them to the ways that social scientists and other professionals read, analyze, and interpret data. In this section, we address some of the key outcomes that students can expect after taking this course.

First, this textbook takes the view that statistics are tools: useful, but not ends in themselves. Thus, we will not take a “mathematical” approach to the subject, although we will cover enough material so that you can develop a basic understanding of why statistics do what they do. The operations in this text can be carried out manually, but since calculators are rarely used for this purpose in real-world settings, we also demonstrate how to carry out most operations using SPSS, a valuable statistical software package. Many professionals use SPSS for the same purposes, but others use Microsoft Excel, SAS, STATA, R (a free and open sourced program), and other software packages.

Second, you will soon become involved in advanced coursework in your major field of study, and you will find that much of the professional literature assumes at least basic statistical literacy. Furthermore, after graduation, many of you will find yourselves in positions—either in a career or in graduate school—where some understanding of statistics will be helpful or perhaps even required. Very few of you will become statisticians, but you must have a grasp of statistics in order to read and critically appreciate the research literature of your discipline. As a student in the social sciences, you simply cannot realize your full potential without a background in statistics.

This textbook is an introduction to statistics as they are utilized in the social sciences and applied in many professional settings. The general objective of the text is to develop an appreciation—a healthy respect—for statistics and their place in the research process. You should emerge from this experience with the ability to use statistics intelligently and to know when other people have done so—and when they have not! Over time, you will become familiar with the advantages and limitations of the more commonly used statistical techniques, and you will learn which operations are appropriate for a given purpose. The overall experience should give you insight into how statistics are used to address social problems, as well as to make informed decisions in most areas of work and life.

## Descriptive and Inferential Statistics

As noted earlier, statistics are tools used to analyze data and answer research questions. Two general classes of statistical techniques, introduced in this section, are available to accomplish this task.

### DESCRIPTIVE STATISTICS

The first class of techniques, known as **descriptive statistics**, is relevant in several different situations:

1. When a researcher needs to summarize or describe the distribution of a single variable. These statistics are called *univariate* (“one variable”) descriptive statistics.
2. When the researcher wishes to describe the relationship between two or more variables. These statistics are called *bivariate* (“two variable”) or *multivariate* (more than two variable) descriptive statistics.



## STATISTICS

## In Everyday Life

## Using Descriptive Statistics

American society is increasingly connected to the Internet. In 2016, about 82% of American households had access to the Internet, a dramatic increase from 53% in 2000. However, connectedness is dependent on social class: 97% of the most affluent Americans used the Internet vs. only 59% of the least affluent. What are the implications of this pattern? Are the less affluent being left behind? Will their lower use of this increasingly essential resource affect the education and job prospects of their children and thus perpetuate their lower income over the generations? What additional information would you need to answer these questions?

Source: U.S. Bureau of the Census. 2018. Computer and Internet Use in the United States, 2016. Available at <https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf>

**Univariate Descriptive Statistics.** Many of these statistics are probably familiar to you. For example, percentages, averages, and graphs can all be used to describe single variables.

To illustrate the usefulness of univariate descriptive statistics, consider the following problem: Suppose you wanted to summarize the income distribution of a community of 10,000 families. How would you do it? Obviously, you couldn't simply list all the incomes and let it go at that. You would want to use some summary measures of the overall distribution—perhaps a graph, an average, or the percentage of incomes that are low, moderate, or high. Whatever method you choose, its function is to reduce these thousands of items of information into a few clear, concise, and easily understood numbers. The process of using a few numbers to summarize many, called **data reduction**, is the basic goal of univariate descriptive statistics. Part I of this text is devoted to these statistics.

**Bivariate and Multivariate Descriptive Statistics.** The second type of descriptive statistics is designed to help us understand the relationship between two or more variables. These statistics, called **measures of association**, allow us to quantify the strength and direction of a relationship. We can use these statistics to investigate two matters of central theoretical and practical importance to any science: causation and prediction. Measures of association help us disentangle the connections between variables and trace the ways in which some variables might affect others. We can also use them to predict scores on one variable from scores on another.

For example, suppose you were interested in the relationship between two variables—amount of study time and grades—and had gathered data from a group of college students. By calculating the appropriate measure of association, you could determine the strength of the relationship and its direction. Suppose you found a strong, positive relationship. This would indicate that “study time” and “grade” were closely related (strength of the relationship) and that as one increased in value, the other also increased (direction of the relationship). You could make predictions from one variable to the other (“the longer the study time, the higher the grade”).

Measures of association can give us valuable information about relationships between variables and help us understand how one variable causes another.



One important point to keep in mind about these statistics, however, is that they cannot, by themselves, *prove* that two variables are causally related. Even if a measure of association shows a very strong relationship between study time and grades, we cannot conclude that one variable causes the other. Correlation is not the same thing as causation, and the mere existence of a correlation cannot prove that a causal relationship exists between variables. We will consider bivariate associations or correlations in Part III of this text, and we will cover multivariate analysis in Part IV.

## INFERENTIAL STATISTICS

This second class of statistical techniques becomes relevant when we wish to generalize to a **population**, the total collection of all cases in which the researcher is interested and wishes to understand better. Examples of possible populations would be voters in the United States, all parliamentary democracies, or all unemployed people in Atlanta.

Populations can theoretically range from enormous (“all humanity”) to quite small (all sophomores on your campus) but are usually fairly large. Social scientists almost never have the resources or time to test every case in a population, hence the need for **inferential statistics**. This class of techniques involves using information from **samples** (carefully chosen subsets of the population) to make inferences about populations. Because they have fewer cases, samples are cheaper to assemble, and—if the proper techniques are followed—generalizations based on samples can be very accurate representations of populations.

Many of the concepts and procedures involved in inferential statistics may be unfamiliar, but most of us are experienced consumers of inferential statistics—most familiarly, perhaps, in the form of public-opinion polls and election projections. When a public-opinion poll reports that 42% of the American electorate plans to vote for a certain presidential candidate, it is essentially reporting a generalization to a population (“the American electorate”—which numbers over 130 million people) from a carefully drawn sample (usually between 1000 and 3000 respondents). Inferential statistics will occupy our attention in Part II of this book. (*For practice in describing different statistical applications, see problems 1.3 and 1.7.*)

### STATISTICS

### In Everyday Life

#### Using Inferential Statistics

In 2019, a sample of adult Americans was asked about same-sex marriage. About 61% were in favor of allowing gays and lesbians to legally marry, a clear majority of the population. Further, surveys show that the level of support has steadily increased since the turn of the century: in 2001, only 35% of the public supported same-sex marriage. Is American society reaching a consensus on this issue? Before coming to a conclusion, remember that the level of support varies dramatically from group to group. For example, in 2019, 79% of Americans with no religious affiliation supported same-sex marriages vs. only 29% of white Evangelical Protestants.

Source: Pew Research Center. 2019. “Religion and Public Life.” Available at <https://www.pewforum.org/fact-sheet/changing-attitudes-on-gay-marriage/>

## Discrete and Continuous Variables

In the next chapter, you will begin to encounter the many statistics available to social scientists. One aspect of using statistics that can be puzzling is deciding when to use which statistic. You will learn specific guidelines as you go along, but we will introduce some basic and general guidelines at this point. The first of these concerns discrete and continuous variables; the second, covered in the next section, concerns level of measurement.

### DISCRETE VARIABLES

A variable is said to be **discrete** if it has a basic unit of measurement that cannot be subdivided. For example, number of people per household is a discrete variable. The basic unit is people, a variable that will always be measured in whole numbers; you'll never find 2.7 people living in a specific household. The scores of discrete variables will be 0, 1, 2, 3, or some other whole integer. Other examples of discrete variables include number of siblings, children, or cars. To measure these variables, we count the number of units (people, cars, siblings) for each case (household, person) and record results in whole numbers.

### CONTINUOUS VARIABLES

A variable is **continuous** if it has scores that can be subdivided infinitely (at least theoretically). One example of a continuous variable is time, which can be measured in minutes, seconds, milliseconds (thousandths of a second), nanoseconds (billionths of a second), or even smaller units. In a sense, when we measure a continuous variable, we are always approximating and rounding off the scores. We could report somebody's time in the 100-yard dash as 10.7 seconds or 10.732451 seconds, but, because time can be infinitely subdivided, we will never be able to report the exact time elapsed. Because we cannot work with infinitely long numbers, we must report the scores on continuous variables as if they were discrete.

The scores of discrete and continuous variables may look the same even though we measure and process them differently. This distinction is one of the most basic in statistics and will be one of the criteria we use to choose among various statistics and graphs. (*For practice in distinguishing between discrete and continuous variables, see problems 1.4 through 1.9.*)

## Level of Measurement

A second basic and very important guideline for the selection of statistics is the **level of measurement**, or the mathematical nature of the variables under consideration. Some variables, such as age and income, have numerical scores (years, dollars) and can be analyzed in a variety of ways, using many different statistics. For example, we could summarize these variables with a mean or average and make statements like "The average income of residents of this city is \$43,000" or "The average age of students on this campus is 19.7."

Other variables, such as gender and zip code, have “scores” that are really just labels, not numbers at all, and we have fewer options for analyzing them. The average would be meaningless as a way of describing these variables. Your personal zip code might *look* like a number, but it is merely an arbitrary label that happens to be expressed in digits. These “numbers” cannot be added or divided, and statistics like the average zip code of a group of people are pointless.

Level of measurement is crucial because our statistical analysis must match the mathematical characteristics of our variables. One of the first steps in any quantitative research project is to determine the level of measurement of our variables, and we will consider the level of measurement of variables throughout this text and each time a statistical technique is introduced.

There are three levels of measurement. In order of increasing sophistication, they are nominal, ordinal, and interval-ratio. Each will be discussed separately.

## THE NOMINAL LEVEL OF MEASUREMENT

Variables measured at the nominal level (such as gender) have non-numerical “scores” or categories. Other examples of variables at this level include zip code, political party affiliation, and religious preference. Statistical analysis with nominal-level variables is limited to comparing the relative sizes of the categories and making statements such as “There are more females than males in this dorm” or “The most common zip code on this campus is 20158.”

Let’s look at some examples of nominal-level variables and take a moment to consider some terminology. Earlier in this chapter, we defined a variable as any trait that can vary from case to case. A variable consists of categories or scores—the traits that can vary. We “measure” a variable (e.g., gender) by classifying a case into one of the categories (e.g., male, female, transgender, other) or assigning a score to it. Several examples of nominal-level variables and their scores or categories are presented in Table 1.1.

Note that there is a number assigned to each of the categories in Table 1.1 (e.g., 1 for Protestant, 2 for Catholic, etc.), as is commonly done in quantitative research, especially when the data are being prepared for computer analysis. Recall that, for nominal-level variables, these “numbers” are just labels and cannot be treated mathematically: They cannot be added, divided, or otherwise manipulated. The categories or scores do not make up a mathematical scale; they are different from each other but not “more or less” or “higher or lower” than each other. Males and females differ in terms of gender, but neither category has more

**TABLE 1.1 Some Examples of Nominal-Level Variables and Their Categories**

Variables →	Gender	Political Party Preference	Religious Preference
Categories →	1 Male	1 Democrat	1 Protestant
	2 Female	2 Republican	2 Catholic
	3 Transgender	3 Other	3 Jewish
	4 Other	4 Independent	4 None
			5 Other

TABLE 1.2 Criteria for Stating the Categories of Variables

The categories of variables must	
1.	Be mutually exclusive (Each case must fit into one and only one category.)
2.	Be exhaustive (There must be a category for every case.)
3.	Include elements that are similar (The cases in each category must be similar to each other.)

TABLE 1.3 Four Scales for Measuring Religious Affiliation

Scale A	Scale B	Scale C	Scale D
Protestant	Protestant	Protestant	Protestant
Episcopalian	Catholic	Non-Protestant	Catholic
Catholic	Jewish		Jewish
Jewish			None
None			Other
Other			

or less gender than the other. In the same way, a zip code of 54398 is different from but not “higher” than a zip code of 13427.

Nominal variables are rudimentary, but there are criteria that we need to observe in order to measure them adequately. These criteria, in fact, apply to variables measured at all levels. They are stated in Table 1.2 and illustrated in Table 1.3. We consider and illustrate the criteria one at a time.

The first criterion (“mutually exclusive”) means that there can be no overlap between categories and no confusion or ambiguity about where a case belongs. To illustrate, consider Scale A in Table 1.3, which violates this principle because of the overlap between the categories “Protestant” and “Episcopalian.”

The second criterion (“exhaustive”) means that there must be a place for every case or score. Scale B in Table 1.3 violates this criterion because there is no category for people who have no religious affiliation and for those who belong to religions other than the three stated. Often, we meet this criterion by including an “Other” category (as in Scale D).

The third criterion (“similar”) means that the categories should include cases that are truly comparable. To put it another way, we should avoid lumping apples with oranges. There are no hard-and-fast guidelines for judging whether a category includes similar items. The researcher must make that decision in terms of the specific purpose of the research, and categories that are too broad for some purposes may be perfectly adequate for others. Scale C in Table 1.3 violates this criterion because it uses a category (Non-Protestant, which would include Catholics, Jews, Buddhists, Atheists, and so forth) that seems too broad for meaningful research.

Scale D in Table 1.3 is probably the most common way religious preference has been measured in North America, and this scale would be adequate in many situations. However, it may be too general for some research purposes. For example, an investigation focused on moral questions such as support for

legal abortion would probably have to distinguish among the various Protestant denominations, and an effort to document religious diversity certainly would need to add categories for Buddhists, Muslims, and numerous other religious faiths.

## THE ORDINAL LEVEL OF MEASUREMENT

Variables measured at the ordinal level are more sophisticated than nominal-level variables. They have scores or categories that can be ranked from high to low, so, in addition to classifying cases into categories, we can describe the categories in terms of “more or less” with respect to each other. Thus, with ordinal-level variables, not only can we say that one case is different from another; we can also say that one case is higher or lower, more or less than another.

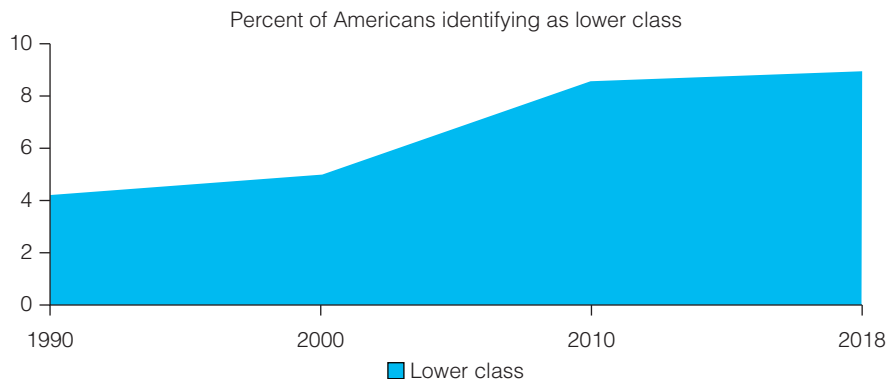
For example, the variable socioeconomic status is usually measured at the ordinal level, often using categories such as those in Table 1.4. Individual cases can be compared in terms of the categories into which they are classified. Thus, an individual classified as a 4 (upper class) would be ranked higher than an

### STATISTICS

### In Everyday Life

#### Changes in Socioeconomic Status in the United States

The poverty rate in the United States has fluctuated between 10% and 15% since the mid-1960s. It normally rises during periods of recession and declines as the economy recovers. It is interesting to note, however, that the percentage of Americans who define themselves as “lower class” is much less than the facts about poverty suggest, although this has been changing over time. According to the General Social Survey, only 4.2% of Americans described themselves as lower class in 1980. This rose to 5% in 1990, 8.6% in 2010, and 9% in 2018. Why does this disparity between self-identification and the realities of poverty persist? What sociological factors might help to explain this phenomenon?



Sources: U.S. Bureau of the Census. Available at <https://www.census.gov/library/stories/2018/09/poverty-rate-drops-third-consecutive-year-2017.html>; National Opinion Research Center. The General Social Survey. Available at <https://gssdataexplorer.norc.umd.edu/>

TABLE 1.4 Measuring Socioeconomic Status

If you were asked to describe yourself, into which of the following classes would you say you belong?	
Score	Class
1	Lower class
2	Working class
3	Middle class
4	Upper class

individual classified as a 2 (working class), and a lower-class person (1) would rank lower than a middle-class person (3). Other variables that are usually measured at the ordinal level include attitude and opinion scales, such as those that measure prejudice, alienation, or political conservatism.

The major limitation of the ordinal level of measurement is that the scores have no absolute or objective meaning: They only represent position with respect to other scores. We can distinguish between high and low scores, but the distance between the scores cannot be described in precise terms. Thus, in terms of the scale shown in Table 1.4, although we know that a social class score of 4 is more than a score of 2, we do not know whether it is twice as much as 2.

Our options for statistical analysis with ordinal-level variables are limited by the fact that we don't know the exact distances from score to score. For example, addition (and other mathematical operations) assumes that the intervals between scores are exactly equal. If the distances from score to score are not equal,  $2 + 2$  might equal 3 or 5 or even 15.

Strictly speaking, statistics such as the average, or mean (which requires that the scores be added together and then divided by the number of scores), are not permitted with ordinal-level variables. The most sophisticated mathematical operation fully justified with an ordinal variable is ranking categories and cases (although, as we will see, it is not unusual for social scientists to violate this rule by calculating an average for an ordinal-level variable).

THE INTERVAL-RATIO LEVEL OF MEASUREMENT<sup>2</sup>

The scores of variables measured at the interval-ratio level are actual numbers that can be analyzed with all possible statistical techniques. This means that we can add or multiply the scores, compute averages or square roots, or perform any other mathematical operation.

There are two crucial differences between ordinal-level and interval-ratio-level variables. First, interval-ratio-level variables have equal intervals from score to score. For example, age is an interval-ratio variable because the unit of measurement (years) has equal intervals (the distance from year to year is

<sup>2</sup>Many statisticians distinguish between the interval level (equal intervals) and the ratio level (true zero point). This text will treat these two levels as one.

365 days). Similarly, if we ask people how many siblings they have, we would produce a variable with equal intervals: Two siblings are 1 more than 1, and 13 is 1 more than 12. Likewise, if I have four children and you have two children, I have twice as many children as you do.

Second, interval-ratio variables have true zero points. That is, the score of 0 for these variables is not arbitrary; it indicates the absence or complete lack of whatever is being measured. For example, the variable “number of siblings” has a true zero point because it is possible to have no siblings at all. Similarly, it is possible to have 0 years of education, no income at all, a score of 0 on a

## ONE STEP AT A TIME

### Determining the Level of Measurement of a Variable

#### Step Operation

1. Inspect the scores or categories of the variable.
2. Can the scores be ranked from low to high?

*Illustration:* Marital status is a variable that may be categorized as follows:

- Married
- Never married
- Divorced
- Widowed

Do people who are widowed have *more* marital status than those who are divorced?

Your answer should be *no*, because categories of marital status cannot be ranked from low to high.

**STOP HERE. The level of measurement is nominal.**

For variables for which the categories *can* be ranked from low to high, proceed to step 3.

3. Is the distance between the scores equal or defined?

*Illustration:* Consider the following scale that measures support for the death penalty.

- Somewhat oppose
- Strongly oppose
- Somewhat support
- Strongly support

People who “strongly” support the death penalty are more in favor than people who “somewhat” support it, but it is unclear from the categories how much additional support to expect. Your answer should be *no*, the distance is not equal or defined.

**STOP HERE. The level of measurement is ordinal.**

For variables for which the scores *are* equally spaced and defined, proceed to step 4.

4. You have determined that the scores or categories of your variable can be ranked from low to high and are equally spaced and defined.

**STOP HERE. The level of measurement is interval-ratio.**

*Illustration:* Consider the following question: What is your current age?

Ages can be ranked from low to high and are defined and equally spaced. Age 4 comes before age 5. The distance between the ages of 4 and 5 is 12 months, just like the distance between the ages of 68 and 69. A person who is 20 has lived exactly twice as long as a person who is 10.

## READING STATISTICS

## An Introduction

Many students are drawn to the social sciences because they want to make a difference in society. It sometimes comes as a shock to them that they need to learn about research methods and statistics. Some students ask, What's the point? One answer to this question is that the analysis of data has become an essential part of everyday decision making in all sectors of society. To learn more about how statistics are used to make evidence-based decisions, see the boxes in this text on Social Research and Data Analysis.

A second answer to this question is that data are analyzed in all forms of published research in the social sciences. If you are interested in becoming a social science researcher, these boxes on Reading Statistics are for you. As you will see, professional researchers use a reporting style that is quite different from the statistical language you will find in this text. Space in reports and research journals can be expensive, and the typical research project requires the analysis of many variables. Thus, a large volume of information must be summarized in few words. Researchers may express in a word or

two a result or an interpretation that will take us a paragraph or more to state.

Because this is an introductory textbook, we have been careful to break down the computation and logic of each calculation and to explain what we are doing when we use statistics. In this text, we will never be concerned with more than a few variables at a time. We will be able to take pages or even entire chapters to develop a statistical idea or concept. Thus, a major purpose of these boxed inserts is to summarize how our comparatively long-winded (but more careful) vocabulary is translated into the concise language of the professional researcher.

When you have difficulty reading words, you consult reference books (especially dictionaries or online search engines) to help you analyze the elements (words) of the passage. When you have difficulty reading statistics, you should do exactly the same thing. We hope you will find this text a valuable reference book, but if you learn enough to be able to use any source to help you read statistics, this text will have fulfilled one of its major goals.

multiple-choice test, and to be 0 years old (although not for very long). Other examples of interval-ratio variables normally include number of children, life expectancy, and years married. All mathematical operations are permitted for variables measured at the interval-ratio level.

## SOME FINAL POINTS

Table 1.5 summarizes the discussion of the three levels of measurement. Note that the number of permitted mathematical operations increases as we move from nominal to ordinal to interval-ratio levels of measurement. Ordinal-level variables are more sophisticated and flexible than nominal-level variables, and interval-ratio-level variables permit the broadest range of mathematical operations.

Let us end this section by making four points. The first stresses the importance of level of measurement, and the next three discuss some common points of confusion in applying this concept.

1. Knowing the level of measurement of a variable is crucial because it tells us which statistics are appropriate and useful. Not all statistics can be used with all variables. As shown in Table 1.5, different statistics require different



**TABLE 1.5 Basic Characteristics of the Three Levels of Measurement**

Levels	Examples	Measurement Procedures	Mathematical Operations Permitted
Nominal	Gender, race, religion, marital status	Classification into categories	Counting number in each category, comparing sizes of categories
Ordinal	Social class, attitude and opinion scales	Classification into categories <i>plus</i> ranking of categories with respect to each other	All of the above <i>plus</i> judgments of “greater than” and “less than”
Interval-Ratio	Age, number of children, income	All of the above <i>plus</i> description of scores in terms of equal units	All of the above <i>plus</i> all other mathematical operations (addition, subtraction, multiplication, division, square roots, etc.)

mathematical operations. For example, computing an average requires addition and division, and finding a median (or middle score) requires that the scores be ranked from high to low. Addition and division are appropriate only for interval-ratio-level variables, and ranking is possible only for variables that are at least ordinal in level of measurement. Your first step in dealing with a variable and selecting appropriate statistics is *always* to determine its level of measurement.

- The distinction made earlier between discrete and continuous variables is a concern only for interval-ratio-level variables. Nominal- and ordinal-level variables are almost always discrete. That is, researchers typically measure these variables by asking respondents to select the single category that best describes them, as illustrated in Table 1.6. The scores of these survey items, as they are stated here, are discrete because respondents must select only one category and because these scores cannot be subdivided.

Interval-ratio variables can be either discrete (number of times you’ve been divorced, which must be a whole integer) or continuous (age, which could be measured to the nanosecond, or even more precisely). Remember that, since we cannot work with infinitely long numbers, continuous interval-ratio variables have to be rounded off and reported as if they were discrete. The distinction between discrete and continuous variables relates more to our options for appropriate statistics or graphs than to the appearance of the variables.

- In determining level of measurement, always examine the way in which the scores of the variable are *actually stated*. This is particularly a problem with variables that are interval-ratio but that have been measured at the ordinal level.

To illustrate, consider income as a variable. If we asked respondents to list their exact income in dollars, we would generate scores that are interval-ratio in level of measurement. Measured in this way, the variable would have a true zero point (no income at all) and equal intervals from score to score (1 dollar). It is more convenient for respondents, however, simply to check the appropriate category from a list of income ranges, as in Table 1.7. The four scores or

**TABLE 1.6 Measuring a Nominal Variable (Marital Status) and an Ordinal Variable (Support for Capital Punishment) as Discrete Variables**

What is your marital status? Are you presently:		Do you support the death penalty for persons convicted of homicide?	
Score	Category	Score	Category
1	Married	1	Strongly support
2	Divorced	2	Somewhat support
3	Separated	3	Neither support nor oppose
4	Widowed	4	Somewhat oppose
5	Single	5	Strongly oppose

**TABLE 1.7 Measuring Income at the Ordinal Level**

Score	Income Range
1	Less than \$25,000
2	\$25,000 to \$49,999
3	\$50,000 to \$99,999
4	\$100,000 or more

categories in Table 1.7 are ordinal in level of measurement because they are unequal in size. It is common for researchers to sacrifice precision (income in actual dollars) for the convenience of the respondents. You should be careful to look at the way in which the variable is measured before making a decision about its level of measurement.

- There is a mismatch between the variables that are usually of most interest to social scientists (race, gender, marital status, attitudes, and opinions) and the most powerful and interesting statistics (such as the mean). The former are typically nominal or ordinal in level of measurement, but the more sophisticated statistics require measurement at the interval-ratio level.

This mismatch creates some real difficulties for social science researchers. On one hand, researchers will want to measure variables at the highest, most precise level of measurement possible. If income is measured in exact dollars, for example, researchers can make precise descriptive statements about the differences between people: “Ms. Parashar earns \$12,547 more than Mr. Wong.” If the same variable is measured in broad, unequal categories, such as those in Table 1.7, comparisons between individuals would be less precise and provide less information: “Ms. Parashar earns more than Mr. Wong.”

On the other hand, given the nature of the disparity, researchers are more likely to treat variables as if they were higher in level of measurement than they actually are. In particular, variables measured at the ordinal level, especially when they have many possible categories or scores, are often treated as if they were interval-ratio and are analyzed with the more powerful, flexible, and interesting statistics available at the higher level. This practice is common, but researchers should be cautious in assessing statistical results and developing interpretations when the level of measurement criterion has been violated.

In conclusion, level of measurement is a very basic characteristic of a variable, and we will always consider it when presenting statistical procedures. Level of measurement is also a major organizing principle for the material that follows, and you should make sure that you are familiar with these guidelines. (*For practice in determining the level of measurement of a variable, see problems 1.4 through 1.9.*)

## SUMMARY

1. The purpose of statistics is to organize, manipulate, and analyze data so that researchers can test theories and answer questions. Along with theory and methodology, statistics are a basic tool used by social scientists to enhance their understanding of the social world.
2. There are two general classes of statistics. Descriptive statistics are used to summarize the distribution of a single variable and the relationships between two or more variables. We use inferential statistics to generalize to populations from random samples.
3. Variables may be either discrete or continuous. They may be measured at any of three different levels. At the nominal level, we can compare category sizes. At the ordinal level, scores can be ranked from high to low. At the interval-ratio level, all mathematical operations are permitted. Interval-ratio-level variables can be either discrete or continuous. Variables at the nominal or ordinal level are almost always discrete.

## GLOSSARY

**Continuous variable.** A variable with a unit of measurement that can be subdivided infinitely.

**Data.** Information expressed as numbers.

**Data reduction.** Summarizing many scores with a few statistics.

**Dependent variable.** A variable that is identified as an effect or outcome. The dependent variable is thought to be caused by the independent variable.

**Descriptive statistics.** The branch of statistics concerned with (1) summarizing the distribution of a single variable or (2) measuring the relationship between two or more variables.

**Discrete variable.** A variable with a basic unit of measurement that cannot be subdivided.

**Hypothesis.** A specific statement, derived from a theory, about the relationship between variables.

**Independent variable.** A variable that is identified as a cause. The independent variable is thought to cause the dependent variable.

**Inferential statistics.** The branch of statistics concerned with making generalizations from samples to populations.

**Level of measurement.** The mathematical characteristic of a variable and the major criterion for selecting

statistical techniques. Variables can be measured at any of three levels, each permitting certain mathematical operations and statistical techniques. The characteristics of the three levels are summarized in Table 1.5.

**Measures of association.** Statistics that summarize the strength and direction of the relationship between variables.

**Population.** The total collection of all cases in which the researcher is interested.

**Quantitative research.** Research projects that collect data or information in the form of numbers.

**Research.** Any process of gathering information systematically and carefully to answer questions or test theories. Statistics are useful for research projects that collect numerical information or data.

**Sample.** A carefully chosen subset of a population. In inferential statistics, information is gathered from a sample and then generalized to a population.

**Statistics.** A set of mathematical techniques for organizing and analyzing data.

**Theory.** A generalized explanation of the relationship between two or more variables.

**Variable.** Any trait that can change values from case to case.

## PROBLEMS

- 1.1** In your own words, describe how statistics might be used to measure changes in divorce rates over time. Using the “wheel of science” as a framework, explain how statistics can be used to link theories about family change with research.
- 1.2** Find a research article in any social science journal. Choose an article on a subject of interest to you, and don’t worry about being able to understand all of the statistics that are reported.
- How much of the article is devoted to statistics?
  - Is the research based on a sample from some population? How large is the sample? How were subjects or cases selected? Can the findings be generalized to some population?
  - What variables are used? Which are independent and which are dependent? For each variable, determine the level of measurement and whether the variable is discrete or continuous.
  - What statistical techniques are used? Try to follow the statistical analysis and see how much you can understand. Save the article and read it again after you finish this course and see if you do any better.
- 1.3** Distinguish between descriptive and inferential statistics. Describe a research situation that would use each type.
- 1.4** For each of the following items from a public-opinion survey, indicate the level of measurement and whether the variable will be discrete or continuous. (*HINT: Remember that only interval-ratio-level variables can be continuous.*)
- In what country were you born? \_\_\_\_\_
  - What is your age? \_\_\_\_\_
  - How many years of school have you completed? \_\_\_\_\_
  - What is your occupation? \_\_\_\_\_
  - If you were asked to use one of these four names for your social class, in which would you say you belonged?  
\_\_\_\_\_ Upper \_\_\_\_\_ Middle  
\_\_\_\_\_ Working \_\_\_\_\_ Lower
  - What is your grade-point average? \_\_\_\_\_
  - What is your major? \_\_\_\_\_
  - The only way to deal with the drug problem is to legalize all drugs.  
\_\_\_\_\_ Strongly disagree  
\_\_\_\_\_ Disagree  
\_\_\_\_\_ Undecided
- \_\_\_\_\_ Agree  
\_\_\_\_\_ Strongly agree
- What is your astrological sign? \_\_\_\_\_
  - How many brothers and sisters do you have? \_\_\_\_\_
- 1.5** Following are brief descriptions of how researchers measured a variable. For each situation, determine the level of measurement of the variable and whether it is continuous or discrete.
- Race or ethnicity.** Check all that apply:  
\_\_\_\_\_ Black or African American  
\_\_\_\_\_ White  
\_\_\_\_\_ Hispanic  
\_\_\_\_\_ Asian  
\_\_\_\_\_ Native Hawaiian or Other Pacific Islander  
\_\_\_\_\_ American Indian or Alaskan Native  
\_\_\_\_\_ Other (Please specify: \_\_\_\_\_)
  - Honesty.** Subjects were observed as they passed by a spot where an apparently lost wallet was lying. The wallet contained money and complete identification. Subjects were classified into one of the following categories:  
\_\_\_\_\_ Returned the wallet with money  
\_\_\_\_\_ Returned the wallet but kept the money  
\_\_\_\_\_ Did not return the wallet
  - Social class.** What was your family situation when you were 16 years old?  
\_\_\_\_\_ Very well off compared to other families  
\_\_\_\_\_ About average  
\_\_\_\_\_ Not so well off
  - Education.** How many years of schooling have you completed? \_\_\_\_\_
  - Racial integration on campus.** Students were observed during lunchtime at the cafeteria for a month. The number of students sitting with students of other races was counted for each meal period.
  - Number of children.** How many children have you ever had? \_\_\_\_\_
  - Student seating patterns in classrooms.** On the first day of class, instructors noted where each student sat. Seating patterns were remeasured every two weeks until the end of the semester. Each student was classified as:  
\_\_\_\_\_ Same seat as at last measurement  
\_\_\_\_\_ Adjacent seat  
\_\_\_\_\_ Different seat, not adjacent  
\_\_\_\_\_ Absent

- h. Physicians per capita.** The number of physicians was counted in each of 50 cities. The researchers used population data to compute the number of physicians per capita.
- i. Physical attractiveness.** A panel of 10 judges rated each of 50 photos of a mixed-race sample of males and females for physical attractiveness on a scale from 0 to 20, with 20 being the highest score.
- j. Number of accidents.** The number of traffic accidents per year for each of 20 intersections was recorded. Also, each accident was rated as:
  - \_\_\_\_\_ Minor damage, no injuries
  - \_\_\_\_\_ Moderate damage, personal injury requiring hospitalization
  - \_\_\_\_\_ Severe damage and injury

**1.6** For each research project listed here, identify the variables and classify them in terms of level of measurement, and whether they are discrete or continuous. Is each variable independent or dependent?

- a.** For a research project in a political science course, a student collected information for 50 nations. She used infant mortality rates as a measure of quality of life and the percentage of all adults who are permitted to vote as a measure of democratization. Her hypothesis was that quality of life is higher in more democratic nations.
- b.** A highway engineer wonders whether increasing the speed limit on a heavily traveled highway will result in more accidents. He plans to collect information on traffic volume, number of accidents, and number of fatalities for the six-month periods before and after the speed limit is changed.
- c.** Students are planning a program to promote “safe sex” and awareness of other health concerns on campus. To measure the effectiveness of the program, they plan to survey students about their knowledge of safe sex practices before and after the program.
- d.** A graduate student asks 500 female students whether they have experienced any sexual harassment on campus. Each student is asked to estimate the frequency of these incidents as either “often, sometimes, rarely, or never.” The researcher also gathers data on age and major to see whether there is any connection between these variables and frequency of sexual harassment.
- e.** A supervisor in the solid waste management division of city government is assessing two different methods of trash collection. One area of the city is served by trucks with two-man crews who do “backyard” pickups, and the rest of the city is

served by “high-tech” single-person trucks with curbside pickup. The assessment measures include the number of complaints received from the two different areas over a six-month period, the amount of time per day required to service each area, and the cost per ton of trash collected.

- f.** Does tolerance for diversity vary by race or ethnicity? Samples of white, black, Asian, Hispanic, and Native Americans have been given a survey that measures their interest in and appreciation of cultures other than their own.
- g.** States have drastically cut their budgets for mental health care. Will this increase the number of homeless people? A researcher contacts a number of agencies serving the homeless in each state and develops an estimate of the size of the homeless population before and after the cuts.
- h.** The adult bookstore near campus has been raided and closed by the police. Your social research class has decided to poll a sample of students to find out: whether he or she supports the store’s closing, how many times each has visited the store, and whether he or she agrees that “pornography causes sexual assaults on women.” The class also collects information on the gender, political philosophy, and major of the students.

**1.7** For each of the following research situations, identify the level of measurement of all variables and indicate whether they are discrete or continuous. Also, decide which statistical applications are used: descriptive statistics (single variable), descriptive statistics (two or more variables), or inferential statistics. Remember that it is quite common for a given situation to require more than one type of application.

- a.** The administration of your university is proposing a change in parking policy. You select a random sample of students and ask each one how strongly they favor or oppose the change. You then use the results to estimate the support for the change in the entire student body.
- b.** You ask everyone in your social research class for their highest letter grade in any math course and their grade (percentage of items correct) on a recent statistics test. You compare the two sets of scores to see whether there is a relationship.
- c.** Your aunt is running for mayor and hires you to question a random sample of voters about their concerns. Specifically, she wants to use this information to characterize the entire city in terms of political party affiliation, gender, and what percentage favor widening of the main street in town.

- d. Several years ago, a state reinstituted the death penalty for first-degree homicide. Did this reduce the homicide rate? A researcher has gathered information on the number of homicides in the state for the two-year periods before and after the change.
  - e. A local automobile dealer is concerned about customer satisfaction. He mails a survey form to all customers for the past year and asks whether they are satisfied, very satisfied, or not satisfied with their purchases.
- 1.8** For each of the first 20 items in the General Social Survey (see Appendix G), indicate the level of measurement and whether the variable is continuous or discrete.
- 1.9** Identify all variables in the research situations listed below and classify them by their level of measurement. Which variables are independent and which are dependent?
- a. A researcher is wondering about racial preferences in dating among college students and asks a large sample of undergraduates about their own racial self-identification. She then asks the respondents to rank some racial-ethnic categories (stated as: white, black, Latino, Asian) in terms of desirability as potential dates.
  - b. For high school students, does GPA affect sexual activity? A sample has been interviewed about the number of different romantic relationships they have had, the number of times they have had sexual intercourse, and their high school grade-point average.
  - c. Several hundred voting precincts across the nation have been classified in terms of percentage of minority voters, voting turnout, and percentage of local elected officials who are members of minority groups. Do precincts with higher percentages of minority voters have lower turnout? Do precincts with higher percentages of minority elected officials have higher turnout?
  - d. As nations become more affluent (as measured by per capita income), does the percentage of children enrolled in school increase? Is this relationship different for boys and girls?
  - e. Does the level of support for gun control vary by number of years of schooling? Does this relationship vary by gender, region of the country, or political party preference? Support for gun control was measured by a five-point scale that ranged from “strongly in favor” to “strongly opposed.”

## YOU ARE THE RESEARCHER

### Introduction

The best way—maybe the only way—to learn statistics and to appreciate their importance is to apply and use them. This includes selecting the correct statistic for a given purpose, doing the calculations, and interpreting the result. We have included extensive end-of-chapter problems to give you multiple opportunities to select and calculate statistics and say what they mean. Many of these problems have been written so that they can be solved with just a simple hand calculator. We’ve purposely kept the number of cases unrealistically low so that the tedium of lengthy calculations would not interfere with the learning process. These problems present an important and useful opportunity to develop your statistical skills.

As important as they are, these end-of-chapter problems are simplified and several steps removed from the complex realities of social science research and data analysis. To provide a more realistic statistical experience, we have included a feature called “You Are the Researcher,” in which you will walk through many of the steps of a research project, making decisions about how to apply your growing knowledge of research and statistics and interpreting the statistical output you generate.

To conduct research and data analysis, you need two things. First, you need something to analyze, a database or a collection of information. These days, databases are often very large and include hundreds or thousands or more of bits of information. They are far too large to analyze with even the most sophisticated hand

calculator, and this brings us to the second thing you need: a computerized statistical analysis program. We address these two needs separately.

In the “You Are the Researcher” exercises, you will usually analyze the 2018 General Social Survey (GSS), although several other databases are supplied with this text. Please visit [www.cengagebrain.com](http://www.cengagebrain.com) to view and download the databases. The GSS is a public-opinion poll that has been conducted on nationally representative samples of citizens of the United States since 1972. The 2018 GSS includes more than 1000 items of information (e.g., gender, race, level of education) and questions covering a broad range of social and political issues (e.g., support for legal abortion, political beliefs and values, and levels of trust in other people). The GSS and the other databases are actual, “real-life” data, so you have the opportunity to practice your statistical skills in a realistic context.

To analyze the GSS—or any large database—you will use a computerized statistical software package. There are several such programs commonly available, such as SAS, STATA, and R. The one featured in this text is called the Statistical Package for the Social Sciences (SPSS). A statistical package is a set of computer programs designed to analyze data. The advantage of these packages is that, because the programs are already written, you can capitalize on the power of the computer with minimal computer literacy and virtually no programming experience. Be sure to read Appendix F before attempting any data analysis.

Some of you may not have access to SPSS on your campus but you can still conduct these research projects by using the student version of SPSS.<sup>3</sup> This student version is limited to smaller available online with fewer than 1500 respondents and 50 variables.

In most of the research exercises, which begin in Chapter 2, you will make the same kinds of decisions as do professional researchers and move through some of the steps of a research project. You will select variables and statistics, generate and analyze output, and express your conclusions. When you finish these exercises, you will be well prepared to conduct your own research project (within limits, of course) and perhaps make a contribution to the ever-growing social science research literature.

---

<sup>3</sup>The student version of SPSS is available online.