SECOND EDITION

APPLIED STATISTICS FOR THE SOCIAL AND HEALTH SCIENCES

RACHEL A. GORDON

ROUTLEDGE

Table of Contents in Brief

Preface xv

Part I
Getting Ready 1

1 Considering Examples of Scholarly Publications Modeling Social and Health Variables 3

2 Planning and Starting a Quantitative Research Project with Existing Data 23

Part II
Describing the Data 87

3 Graphing and Summarizing Individual Variables 89

4 Introducing Population Estimation and Hypothesis Testing 151

5 Estimating and Testing the Association between Two Variables 211

Part III
Estimating and Presenting Linear Regression Models 251

6 Introducing the Linear Regression Model with Two Continuous Variables 253

7 Considering Nonlinearity and Nonconstant Variance 321

8 Including Categorical Predictor Variables 407

9 Including More Than One Predictor Variable in the Model 449

10 Considering Interactions among Predictor Variables 525

Part IV
Estimating and Presenting Generalized Linear Models 595

11 Introducing the Generalized Linear Regression Model 597

12 Analyzing Dichotomous Outcomes 631

13 Analyzing Multi-Category Outcomes and Offering a Roadmap to Additional Models 705

Glossary/Index 751

Table of Contents in Detail

Preface xv

Part I
Getting Ready 1

1 Considering Examples of Scholarly Publications Modeling Social and Health Variables 3
- 1.1 What Is Regression Analysis? 4
- 1.2 Critiques, Debates, and Tensions 7
- 1.3 Literature Excerpt 1.1 9
- 1.4 Literature Excerpt 1.2 10
- 1.5 Literature Excerpt 1.3 14
- 1.6 Summary 20

2 Planning and Starting a Quantitative Research Project with Existing Data 23
- 2.1 Locating and Choosing Secondary Datasets 24
- 2.2 Starting the Quantitative Project 36
- 2.3 Programming for Transparency, Portability, and Reproducibility 67
- 2.4 A Cautionary Tale Encouraging Careful Checking and Coding 80
- 2.5 Summary 81

Part II
Describing the Data 87

3 Graphing and Summarizing Individual Variables 89
- 3.1 Types of Variables 91
- 3.2 Literature Excerpts 3.1 and 3.2 97
- 3.3 Nominal Variables 100
- 3.4 Ordinal Variables 107
- 3.5 Interval/Ratio Variables 112
- 3.6 Weighted Statistics 126
- 3.7 Creating a Descriptive Table 142

4 Introducing Population Estimation and Hypothesis Testing 151
- 4.1 Statistical Inference 153

4.2 Population and Sample Distributions 157
4.3 The Sampling Distribution 167
4.4 Statistical Inference 175
4.5 Other Common Theoretical Distributions 197

5 Estimating and Testing the Association between Two Variables 211
5.1 Literature Excerpts 212
5.2 One Categorical and One Interval Variable 216
5.3 Two Categorical Variables 228
5.4 Two Interval Variables 237
5.5 Weighted Statistics 243
5.6 Automating Table Creation 246

Part III
Estimating and Presenting Linear Regression Models 251

6 Introducing the Linear Regression Model with Two Continuous Variables 253
Chapter Examples 254
Revisiting Literature Excerpt 1.1 255
6.1 Algebraic and Geometric Representations of Bivariate Regression 256
6.2 The Population Regression Line 262
6.3 The Sample Regression Line 265
6.4 Ordinary Least Squares Estimators 268
6.5 Complex Sampling Designs 316

7 Considering Nonlinearity and Nonconstant Variance 321
7.1 A Note about Judgment and Jargon 322
7.2 Nonlinear Relationships 322
7.3 High Leverage, Conditionally Outlying, and Influential Observations 382
7.4 Heteroskedasticity 399

8 Including Categorical Predictor Variables 407
8.1 Why Is a Different Approach Needed for Categorical Predictor Variables? 408
8.2 How Do We Define Indicator Variables? 409
8.3 Interpreting Coefficients of Indicator Terms in Regression Models 420
8.4 Putting It All Together 442

9 Including More Than One Predictor Variable in the Model 449
9.1 Algebraic and Geometric Representations of Multiple Regression 450
9.2 OLS Estimation of the Multiple Regression Model 456
9.3 Conducting Multiple Hypothesis Tests 467
9.4 General Linear *F*-Test 471
9.5 Information Criteria 488
9.6 Literature Excerpt 9.1 490

9.7 The Strengths and Limits of Adding Predictor Variables to the Model 492
9.8 Literature Excerpt 9.2 493
9.9 Defining Confounders, Mediators, and Suppressor Variables 495
9.10 Omitted Variable Bias 509
9.11 Multicollinearity 511

10 Considering Interactions among Predictor Variables 525
10.1 Literature Excerpt 10.1 526
10.2 Interactions between Two Categorical Variables 528
10.3 Interactions between a Categorical and a Continuous Variable 549
10.4 Chow Test 564
10.5 Interactions between Two Continuous Variables 579
10.6 Literature Excerpt 10.2 589

Part IV
Estimating and Presenting Generalized Linear Models 595

11 Introducing the Generalized Linear Regression Model 597
11.1 Literature Excerpt 598
11.2 Maximum Likelihood Estimation 600
11.3 Hypothesis Testing with Maximum Likelihood Estimation 615
11.4 The Generalized Linear Model 620

12 Analyzing Dichotomous Outcomes 631
12.1 Literature Excerpt 12.1 632
12.2 Linear Probability Model 636
12.3 Generalized Linear Model 644
12.4 Goodness of Fit 658
12.5 Interpretation 676

13 Analyzing Multi-Category Outcomes and Offering a Roadmap to Additional Models 705
13.1 Ordered Logit 705
13.2 Multinomial Logit 715
13.3 Empirical Example 727
13.4 A Roadmap to Statistical Methods 743
13.5 A Roadmap to Locating Courses and Resources 747

Glossary/Index 751

SOCIAL SCIENCE / STATISTICS

"This book is a teacher's dream. Not only does it provide a comprehensive discussion of statistics as it is actually practiced in the social and health sciences, it also provides detailed guidance on how to carry out such analyses using Stata, one of the best available and widely used statistical packages. Finally it provides numerous examples drawn directly from the research literature. I know of no other book like it."
Richard Campbell, ***Emeritus Professor of Public Health, University of Illinois at Chicago, USA***

"I taught graduate level statistics for more than 40 years. I always pieced together material from several different textbooks, software manuals, and published articles, since no one volume met the need for entering graduate students with appropriate content coverage. The second edition of Rachel Gordon's book meets these needs better than any other volume I have seen."
George Farkas, ***Distinguished Emeritus Professor of Education, University of California, Irvine, USA***

"I have used the first edition in my multidisciplinary graduate-level statistics course for five years. Gordon's ability to translate complex information into practical, realworld examples that are applicable and engaging for students across the social sciences and health disciplines has helped her textbook stand out. The second edition enhances this even further, bringing the material fully up-to-date with recent advances, and displaying a much-needed focus on developing students' coding skills as well as their statistical knowledge. I anticipate this becoming a standard textbook in the field for years to come."
Jeffrey E. Stokes, ***Assistant Professor of Gerontology & Undergraduate Director of Aging Studies Program, University of Massachusetts Boston, USA***

"I have used—and loved—the first edition of this book for nearly a decade. However, I was thrilled to see that the new edition promises to retain the rigor and clarity of purpose of the first edition, but in a more focused and streamlined package. I look forward to adopting this book for my courses for the next decade and beyond."
Jeffrey M. Timberlake, ***Professor of Sociology, University of Cincinnati, USA***

Covering basic univariate and bivariate statistics and regression models for nominal, ordinal, and interval outcomes, *Applied Statistics for the Social and Health Sciences* provides graduate students in the social and health sciences with fundamental skills to estimate, interpret, and publish quantitative research using contemporary standards.

Reflecting the growing importance of "Big Data" in the social and health sciences, this thoroughly revised and streamlined new edition covers best practice in the use of statistics in social and health sciences, draws upon new literatures and empirical examples, and highlights the importance of statistical programming, including coding, reproducibility, transparency, and open science.

Key features of the book include:

- interweaving the teaching of statistical concepts with examples from publicly available social and health science data and literature excerpts;
- thoroughly integrating the teaching of statistical theory with the teaching of data access, processing, and analysis in Stata;
- recognizing debates and critiques of the origins and uses of quantitative methods.

Rachel A. Gordon is Associate Dean for Research and Administration and Professor of Health Studies in the College of Health and Human Sciences at Northern Illinois University, USA.

an **informa** business

Routledge
Taylor & Francis Group
www.routledge.com

Routledge titles are available as eBook editions in a range of digital formats

ISBN 978-1-032-32344-2