Applied Social Science Methodology
An Introductory Guide

This textbook provides a clear, concise, and comprehensive introduction to methodological issues encountered by the various social science disciplines. It emphasizes applications, with detailed examples, so that readers can put these methods to work in their research. Within a unified framework, John Gerring and Dino Christenson integrate a variety of methods – descriptive and causal, observational and experimental, qualitative and quantitative. The text covers a wide range of topics including research design, data-gathering techniques, statistics, theoretical frameworks, and social science writing. It is designed both for those attempting to make sense of social science, as well as those aiming to conduct original research. The text is complemented by practice questions, exercises, examples, key term highlighting, and additional resources, including related readings and websites. An essential resource for undergraduate and postgraduate programs in communications, criminal justice, economics, business, finance, management, education, environmental policy, international development, law, political science, public health, public policy, social work, sociology, and urban planning.


DINO P. CHRISTENSON is Associate Professor of Political Science at Boston University and a Faculty Affiliate at the Hariri Institute for Computational Science & Engineering. His articles have appeared in American Political Science Review, American Journal of Political Science, Political Behavior, and Social Networks, among others.
Applied Social Science Methodology
An Introductory Guide

JOHN GERRING
University of Texas at Austin

DINO CHRISTENSON
Boston University
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>xvi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xviii</td>
</tr>
<tr>
<td>Abbreviations and Notation</td>
<td>xx</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xxiii</td>
</tr>
<tr>
<td>Preface</td>
<td>xxv</td>
</tr>
</tbody>
</table>

## PART I  BUILDING BLOCKS
1. A Unified Framework  
2. Arguments  
3. Concepts and Measures  
4. Analyses  

## PART II  CAUSALITY
5. Causal Frameworks  
6. Causal Hypotheses and Analyses  
7. Experimental Designs  
8. Large-N Observational Designs  
9. Case Study Designs  
10. Diverse Tools of Causal Inference  

## PART III  PROCESS AND PRESENTATION
11. Reading and Reviewing  
12. Brainstorming  
13. Data Gathering  
14. Writing  
15. Speaking  
16. Ethics  

## PART IV  STATISTICS
17. Data Management  
18. Univariate Statistics  
19. Probability Distributions  
20. Statistical Inference  

## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Bivariate Statistics</td>
<td>313</td>
</tr>
<tr>
<td>22</td>
<td>Regression</td>
<td>331</td>
</tr>
<tr>
<td>23</td>
<td>Causal Inference</td>
<td>353</td>
</tr>
<tr>
<td></td>
<td><strong>Appendix</strong></td>
<td>371</td>
</tr>
<tr>
<td></td>
<td><strong>Notes</strong></td>
<td>375</td>
</tr>
<tr>
<td></td>
<td><strong>References</strong></td>
<td>383</td>
</tr>
<tr>
<td></td>
<td><strong>Indexes</strong></td>
<td>397</td>
</tr>
</tbody>
</table>
## Detailed Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>xvi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xviii</td>
</tr>
<tr>
<td>Abbreviations and Notation</td>
<td>xx</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xxiii</td>
</tr>
<tr>
<td>Preface</td>
<td>xxv</td>
</tr>
<tr>
<td><strong>PART I BUILDING BLOCKS</strong></td>
<td></td>
</tr>
<tr>
<td>1 A Unified Framework</td>
<td>1</td>
</tr>
<tr>
<td>The Purpose of Unity</td>
<td>3</td>
</tr>
<tr>
<td>Examples</td>
<td>5</td>
</tr>
<tr>
<td>Worker-Training Programs</td>
<td>7</td>
</tr>
<tr>
<td>Social Capital</td>
<td>8</td>
</tr>
<tr>
<td>Democracy</td>
<td>9</td>
</tr>
<tr>
<td>Conclusions</td>
<td>11</td>
</tr>
<tr>
<td>Key Terms</td>
<td>13</td>
</tr>
<tr>
<td><strong>2 Arguments</strong></td>
<td></td>
</tr>
<tr>
<td>Descriptive Arguments</td>
<td>14</td>
</tr>
<tr>
<td>Associations</td>
<td>14</td>
</tr>
<tr>
<td>Syntheses</td>
<td>15</td>
</tr>
<tr>
<td>Simple Typologies</td>
<td>16</td>
</tr>
<tr>
<td>Periodization</td>
<td>17</td>
</tr>
<tr>
<td>Matrix Typologies</td>
<td>18</td>
</tr>
<tr>
<td>Taxonomies</td>
<td>19</td>
</tr>
<tr>
<td>Causal Arguments</td>
<td>19</td>
</tr>
<tr>
<td>Other Arguments</td>
<td>21</td>
</tr>
<tr>
<td>Good Arguments</td>
<td>22</td>
</tr>
<tr>
<td>Precision</td>
<td>24</td>
</tr>
<tr>
<td>Generality</td>
<td>25</td>
</tr>
<tr>
<td>Boundedness</td>
<td>25</td>
</tr>
<tr>
<td>Parsimony</td>
<td>26</td>
</tr>
<tr>
<td>Logical Coherence</td>
<td>26</td>
</tr>
<tr>
<td>Commensurability</td>
<td>27</td>
</tr>
<tr>
<td>Innovation</td>
<td>27</td>
</tr>
<tr>
<td>Relevance</td>
<td>28</td>
</tr>
</tbody>
</table>
Detailed Contents

Conclusions 29
Key Terms 30

3 Concepts and Measures 31
Concepts 31
Resonance 32
Internal Coherence 32
External Differentiation 33
Theoretical Utility 34
Consistency 34

Strategies of Concept Formation 35
Minimal 35
Maximal 37
Coda 38

Measures 38
Levels of Abstraction 39
Scales 40
Aggregation 42
Objectives 45

Conclusions 46
Key Terms 46

4 Analyses 47
Definitions 47
Precision and Validity 50
Internal and External Validity 53
Sample Representativeness 54
Sample Size (N) 55
Probability Sampling 56
Non-Probability Sampling 57
Missing-ness 58

Conclusions 60
Key Terms 60

PART II CAUSALITY 63

5 Causal Frameworks 65
Motivational Frameworks 66
Interests 66
Norms 67
Psychology 68
Structural Frameworks 69
Material Factors 69
## Detailed Contents

- **Human Capital/Demography**
  - 70
- **Institutions**
  - 71
- **Interactive Frameworks**
  - 72
  - **Adaptation**
  - 72
  - **Coordination**
  - 73
  - **Diffusion**
  - 74
  - **Networks**
  - 75
  - **Path Dependence**
  - 77
- **Building a Theory**
  - 78
  - **An Example: Birth**
  - 79
  - **Formal Models**
  - 80
- **Conclusions**
  - 83
- **Key Terms**
  - 84

6 **Causal Hypotheses and Analyses**

- **Causality**
  - 85
- **Criteria of a Causal Hypothesis**
  - 89
  - **Clarity**
  - 91
  - **Manipulability**
  - 91
  - **Precedence**
  - 92
  - **Impact**
  - 94
  - **Mechanism**
  - 94
- **Causal Analysis**
  - 95
  - **Covariation**
  - 95
  - **Comparability**
  - 98
- **Conclusions**
  - 99
- **Key Terms**
  - 99

7 **Experimental Designs**

- **Experiments With and Without Confounding**
  - 101
  - **Non-Compliance**
  - 104
  - **Contamination**
  - 105
  - **Compound Treatments**
  - 106
- **Varieties of Experiments**
  - 107
  - **Typology of Designs**
  - 108
  - **Randomization Mechanisms**
  - 110
  - **Research Settings**
  - 110
- **Examples**
  - 111
  - **Employment Discrimination**
  - 111
  - **Corruption Control**
  - 112
  - **Historic Election Campaigns**
  - 113
  - **Gender and Leadership**
  - 113
8 Large-N Observational Designs

Cross-Sectional Designs
  Example
  Time-Series Designs
  Example
  Time-Series Cross-Section (TSCS) Designs
  Example
  Regression-Discontinuity (RD) Designs
  Examples
  Instrumental-Variable (IV) Designs
  Example
  Conclusions
  Key Terms

9 Case Study Designs

Case Selection
  Exploratory
  Diagnostic
Omnibus Criteria
  Intrinsic Importance
  Data Availability
  Logistics
  Case Independence
  Sample Representativeness
Within-Case Analysis
  Identifying a Hypothesis/Theory
  Testing Theories
  Temporality
  Examining Background Assumptions
Strengths and Weaknesses
  Key Terms

10 Diverse Tools of Causal Inference

Identifying and Avoiding Confounders
Research Designs
And the Winner Is . . .
Multiple Methods
Conclusions
Key Terms
## Detailed Contents

### PART III PROCESS AND PRESENTATION

#### 11 Reading and Reviewing

- Identifying Social Science Sources
  - Journals
  - Books
  - Working Papers and Blogs
  - General Observations
- Surveying the Literature
  - Recent Publications
  - Search Terms
  - Citation Searches
  - Other Venues
  - General Observations
- Reading Strategically
- Reading Critically
- Figuring It Out
- Reviewing the Literature
- Taking Notes
- Conclusions

#### 12 Brainstorming

- Study the Tradition
- Begin Where You Are
- Get Off Your Home Turf
- Play With Ideas
- Practice Disbelief
- Observe Empathically
- Theorize Wildly
- Think Ahead
- Conclusions

#### 13 Data Gathering

- Standardized Surveys
  - Medium, Location
  - Respondents
  - Compensation
  - Introductory Statement
  - Questionnaire
  - Response-Categories
Detailed Contents

Knowledge, Certainty, and Salience 201
Sensitive Subjects 202
Longitudinal Data 203
Summary 204
Less Structured Interpersonal Settings 205
Interviews 206
Focus Groups 212
Ethnography 214
Unobtrusive Measures 217
Surreptitious Measures 218
Ex Post Measures 219
Data Assessment 221
Divergent Sources 223
Replicability 226
Validity Tests 227
Conclusions 228
Key Terms 228

14 Writing 230
Genres 231
Organization 232
Introduction 234
Literature Review 235
Thesis 237
Method 237
Evidence and Supporting Arguments 237
Conclusion 238
End Matter 238
Variations 239
Style 239
Rules 240
Sources 243
Quotation Formats 245
Citation Formats 246
Editing 247
Conclusions 248
Key Terms 249

15 Speaking 250
Your Public Image 251
Format 251
Visual Aids 252
Practice 253
Detailed Contents

Know Your Audience 253
Speak Naturally (and, if Possible, Grammatically) 254
Relate to the Whole Room 254
Entertain (a Little) 254
Calm Your Nerves 255
Foster Exchange 256
Be Gracious 257
Audience Participation 258
Conclusions 260

16 Ethics 262
Juridical Ethics 262
Protection of Human Subjects 263
Broader Ethical Issues 266
   The Purview of Social Science 267
Conclusions 268
Key Terms 268

PART IV STATISTICS 269
17 Data Management 271
   Qualitative Data 271
   Medium-N Data 272
   Large-N Data 273
   Textual Data 277
   Examples, Data, and Software 279
Conclusions 281
Key Terms 281
18 Univariate Statistics 282
   Frequency Distributions, Bar Graphs, and Histograms 282
   Central Tendency 285
   Variance 287
   Example: Average State Incomes 288
Conclusions 290
Key Terms 291
19 Probability Distributions 292
   Probabilities 292
   Probability Distributions 293
   The Normal Curve 295
   Example: Feelings about Political Institutions 299
Conclusions 300
Key Terms 301
Detailed Contents

20 Statistical Inference 302
   Sampling 302
   Samples and Populations 304
   Hypothesis Testing 306
   Estimating Population Parameters 309
   Example: Average Number of Parties in Democracies 311
   Conclusions 312
   Key Terms 312

21 Bivariate Statistics 313
   Revisiting Levels of Measurement 313
   Cross-Tabulations 314
   Difference of Means 316
   Parametric Models 320
   Example: Group Membership and Ideology 320
   Correlation 322
   Example: Issue Dimensions and Parties 326
   Conclusions 329
   Key Terms 329

22 Regression 331
   Bivariate Regression 331
   Example: Education and Income 339
   Multivariate Regression 343
   Assumptions 345
   Example: Turnout by Region and Income 349
   Conclusions 352
   Key Terms 352

23 Causal Inference 353
   Assumptions and Assignment 353
   Potential Outcomes 355
   Ignorability in Observational Studies 357
   Regression as a Causal Model 359
   Matching as a Causal Model 361
   Example: Worker-Training Programs 366
   Conclusions 369
   Key Terms 370

Appendix 371
   Normal Curve 371
   T-Distribution 373
### Detailed Contents

- **Notes**  
- **References**  
- **Author Index**  
- **Subject Index**
Figures

2.1 Taxonomy of descriptive arguments  
2.2 Possible relationships among two factors  
2.3 Regime taxonomy in tree-diagram format  
4.1 Time-series cross-section dataset  
4.2 Precision and validity  
5.1 Prisoner’s dilemma  
5.2 A Facebook social network  
5.3 A spatial model of vote choice and party positioning  
6.1 A simple causal graph  
6.2 Covariational patterns  
7.1 Experimental data without confounders: an illustration  
7.2 Experimental data without confounders: a causal graph  
7.3 Experimental data without confounders: X/Y plot  
7.4 Experimental data with non-compliance: an illustration  
7.5 Experimental data with contamination: an illustration  
7.6 Causal graph of compound treatment confounding  
8.1 Cross-sectional data: a typical scenario  
8.2 Causal graph with common-cause confounder  
8.3 Conditioning on a common-cause confounder: an illustration  
8.4 Causal graph with mechanistic (post-treatment) confounder  
8.5 Causal graph with collider confounder  
8.6 Causal graph with circular confounding  
8.7 Time-series data: a typical scenario  
8.8 Instrumental-variable (IV) design  
10.1 Typology of confounders using causal graphs  
10.2 Taxonomy of causal research designs  
18.1 Bar graph of party identification  
18.2 Histograms of feelings toward Obama  
18.3 Feeling thermometer survey instrument  
18.4 Common distribution shapes  
18.5 Unimodal distributions and central tendencies  
18.6 Distributions with different variances  
18.7 Comparing distributions with different standard deviations  
19.1 Frequency and probability outcomes: bar graphs of dice rolls  
19.2 The normal curve  
19.3 Interpreting z-scores
List of Figures

20.1  Distributions of sample means                      305
20.2  Varying alpha                                       307
20.3  Hypothesis test results against real state of null hypothesis  309
20.4  T-distribution at different degrees of freedom      310
21.1  Histogram with polygon of differences between means     318
21.2  Scatter plots and direction of correlation            323
21.3  Scatter plots and strength of correlation             323
21.4  Scatter plot of parties and issues                   328
22.1  Fitting lines through a scatter plot                  332
22.2  Residuals from best fitting line in a scatter plot    332
22.3  Regression line                                      333
22.4  Extrapolation and interpolation                       336
22.5  Plot regression line for education and income         341
22.6  Nonlinear relationship                                346
22.7  Additive and interactive relationships                347
22.8  Venn diagrams by degree of collinearity               348
23.1  Covariate distributions for treated and control groups by overlap  364
23.2  Covariate histograms for treated and control groups by sample  364
# Tables

2.1 A matrix typology: regime-types page 19  
2.2 Regime taxonomy in tabular format 20  
2.3 Typology of arguments 23  
2.4 Theorizing: general criteria 25  
3.1 Criteria of concept formation 32  
3.2 Democracy: fundamental attributes 36  
3.3 Measurement strategies 39  
3.4 Typology of scales 41  
3.5 A single index with multiple interpretations 42  
5.1 Causal frameworks 65  
6.1 Causal hypotheses: criteria 90  
6.2 Causal research designs: criteria 96  
7.1 A typology of experimental designs 108  
8.1 Large-\(N\) observational research designs 118  
8.2 Cross-sectional design 120  
8.3 Time-series research designs 125  
8.4 Time-series cross-section (TSCS) designs 129  
8.5 Regression-discontinuity (RD) design 132  
9.1 Case selection strategies 141  
9.2 Case study and large-\(N\) research designs 153  
14.1 Cross-national studies of development and democracy 236  
17.1 Truth-table with “raw” coding 273  
17.2 Truth-table with binary coding 274  
17.3 Reduced truth-table with primitive case types 274  
17.4 Large-\(N\) dataset structure 276  
18.1 Frequency distribution of party identification 283  
18.2 Calculating central tendency and variance 289  
21.1 Hypothesis tests guide 314  
21.2 Crosstab of partisanship by gender 315  
21.3 Distribution of differences between means 318  
21.4 Summary statistics of group membership by ideology 321  
21.5 Issue dimensions and parties data 327  
21.6 Calculating \(r\) for issue dimensions and parties 327  
22.1 Calculating regression for income and education 339  
22.2 Calculating the sum of squared errors 341  
22.3 Explaining income with education 344
List of Tables

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.4</td>
<td>Voter turnout by South and non-South regions</td>
<td>350</td>
</tr>
<tr>
<td>22.5</td>
<td>Voter turnout by region and income</td>
<td>351</td>
</tr>
<tr>
<td>23.1</td>
<td>Estimating treatment effects with both outcomes</td>
<td>356</td>
</tr>
<tr>
<td>23.2</td>
<td>Estimating treatment effects with one outcome</td>
<td>356</td>
</tr>
<tr>
<td>23.3</td>
<td>Original sample worker-training data</td>
<td>366</td>
</tr>
<tr>
<td>23.4</td>
<td>Matched sample worker-training data</td>
<td>367</td>
</tr>
<tr>
<td>23.5</td>
<td>Regression with matched sample</td>
<td>369</td>
</tr>
<tr>
<td>A1</td>
<td>Proportions of area under the normal curve</td>
<td>371</td>
</tr>
<tr>
<td>A2</td>
<td>Critical values of the T-distribution</td>
<td>373</td>
</tr>
</tbody>
</table>
## Abbreviations and Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Event or option</td>
</tr>
<tr>
<td>a</td>
<td>Y-intercept or constant of a regression equation</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Significance level (read as &quot;alpha&quot;)</td>
</tr>
<tr>
<td>$</td>
<td>$</td>
</tr>
<tr>
<td>$\text{ATE}$</td>
<td>Average treatment effect</td>
</tr>
<tr>
<td>$\text{ATT}$</td>
<td>Average treatment effect on the treated</td>
</tr>
<tr>
<td>B</td>
<td>Event or option</td>
</tr>
<tr>
<td>$b$</td>
<td>Slope coefficient for independent variable (read as &quot;beta&quot;)</td>
</tr>
<tr>
<td></td>
<td>Expression of conditionality</td>
</tr>
<tr>
<td>$\text{CI}$</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>$\text{COV}$</td>
<td>Covariance</td>
</tr>
<tr>
<td>$D_{ij}$</td>
<td>Distance from observation $i$ to $j$</td>
</tr>
<tr>
<td>$D_{\text{Greedy}}$</td>
<td>Distance from a greedy algorithm</td>
</tr>
<tr>
<td>$D_{\text{Optimal}}$</td>
<td>Distance from an optimal algorithm</td>
</tr>
<tr>
<td>$\Delta X$</td>
<td>Change in values of $X$ (read as &quot;delta $X$&quot;)</td>
</tr>
<tr>
<td>$\Delta_i$</td>
<td>Causal treatment effect for individual $i$</td>
</tr>
<tr>
<td>$df$</td>
<td>Degrees of freedom</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>Expression of omission of values in a repeated operation (read as &quot;ellipsis&quot;)</td>
</tr>
<tr>
<td>e</td>
<td>Error term or random component, usually of a regression equation</td>
</tr>
<tr>
<td>$E$</td>
<td>Expectation or expected value</td>
</tr>
<tr>
<td>$\text{ESS}_{Y}$</td>
<td>Explained sum of squares for $Y$</td>
</tr>
<tr>
<td>$\exp$</td>
<td>Natural exponential function</td>
</tr>
<tr>
<td>$H_X$</td>
<td>Hypothesis about the effect of $X$ on $Y$</td>
</tr>
<tr>
<td>$i$</td>
<td>Individual or unit in the sample of observations, not $j$</td>
</tr>
<tr>
<td>$\parallel$</td>
<td>Expression of independence</td>
</tr>
<tr>
<td>$j$</td>
<td>Individual or unit in the sample of observations, not $i$</td>
</tr>
<tr>
<td>$k$</td>
<td>Maximum number in a series of variables or coefficients</td>
</tr>
<tr>
<td>$M$</td>
<td>Mechanism or pathway connecting $X$ to $Y$</td>
</tr>
<tr>
<td>$\text{MOE}$</td>
<td>Margin of error</td>
</tr>
<tr>
<td>$MSE_e$</td>
<td>Mean squared errors</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Mean of a population or probability distribution (read as &quot;mu&quot;)</td>
</tr>
<tr>
<td>$N$</td>
<td>Sample size or number of observations, but occasionally units or cases</td>
</tr>
<tr>
<td>$N_{\text{Col}}$</td>
<td>Number of cases in a column</td>
</tr>
</tbody>
</table>
Abbreviations and Notation

(cont.)

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{Row}}$</td>
<td>Number of cases in a row</td>
</tr>
<tr>
<td>$\text{NA}$</td>
<td>Expression that value is not available, unknown or missing</td>
</tr>
<tr>
<td>$1 - \alpha$</td>
<td>Confidence level</td>
</tr>
<tr>
<td>$P$</td>
<td>Probability</td>
</tr>
<tr>
<td>$\rho$</td>
<td>P-value</td>
</tr>
<tr>
<td>$\text{Pct}_{\text{Col}}$</td>
<td>Column percents</td>
</tr>
<tr>
<td>$\text{Pct}_{\text{Row}}$</td>
<td>Row percents</td>
</tr>
<tr>
<td>$\pi$</td>
<td>Constant value, approximately 3.14 (read as “(\pi)”</td>
</tr>
<tr>
<td>$\text{Prop}$</td>
<td>Proportion</td>
</tr>
<tr>
<td>$Q$</td>
<td>Variable; the antecedent cause (to (X)) that may be used as an instrumental variable</td>
</tr>
<tr>
<td>$r$</td>
<td>Pearson’s correlation coefficient</td>
</tr>
<tr>
<td>$r^2$</td>
<td>Coefficient of determination</td>
</tr>
<tr>
<td>$\text{RSS}_e$</td>
<td>Residual sum of squared errors</td>
</tr>
<tr>
<td>$s_{\mu - \mu_t}$</td>
<td>Standard error of difference between means</td>
</tr>
<tr>
<td>$s_b$</td>
<td>Standard error of (b)</td>
</tr>
<tr>
<td>$s^2$</td>
<td>Variance of a sample</td>
</tr>
<tr>
<td>$\text{SE}_e$</td>
<td>Standard error of the estimate</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>Summation operator</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Standard deviation of a population or probability distribution</td>
</tr>
<tr>
<td>$\sigma_{\mu_t}$</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>Variance of a population or probability distribution</td>
</tr>
<tr>
<td>$\text{SP}$</td>
<td>Sum of products</td>
</tr>
<tr>
<td>$SS_X$</td>
<td>Sum of squares for (X)</td>
</tr>
<tr>
<td>$SS_Y$</td>
<td>Sum of squares for (Y)</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Treatment condition for individual (i)</td>
</tr>
<tr>
<td>$\text{Time}_{1-N}$</td>
<td>Time-periods, usually referring to occasions when key variables are measured</td>
</tr>
<tr>
<td>$t$</td>
<td>T-ratio</td>
</tr>
<tr>
<td>(X)</td>
<td>Variable; usually an independent variable of causal interest</td>
</tr>
<tr>
<td>$\overline{X}$</td>
<td>Mean of (X) (read as “(X)-bar”)</td>
</tr>
<tr>
<td>$X_C$</td>
<td>Control group condition, (X = 0) when binary</td>
</tr>
<tr>
<td>$X_T$</td>
<td>Treatment group condition, (X = 1) when binary</td>
</tr>
<tr>
<td>$X \rightarrow Y$</td>
<td>Expression that (X) causes (Y) and (Y) causes (X)</td>
</tr>
<tr>
<td>$X \rightarrow Y$</td>
<td>Expression that (X) causes (Y)</td>
</tr>
<tr>
<td>$X \rightarrow Y$</td>
<td>Expression that (X) covaries with (Y)</td>
</tr>
<tr>
<td>(Y)</td>
<td>Variable; usually the dependent variable or outcome</td>
</tr>
<tr>
<td>$\overline{Y}$</td>
<td>Mean of (Y) (read as “(Y)-bar”)</td>
</tr>
<tr>
<td>$\hat{Y}$</td>
<td>Predicted, fitted or estimated values of (Y) (read as “(Y)-hat”)</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>Outcome or effect for individual (i)</td>
</tr>
</tbody>
</table>
### Abbreviations and Notation

(Cont.)

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{iC}$</td>
<td>Potential outcome if $i$ does not receive treatment (i.e., in control group)</td>
</tr>
<tr>
<td>$Y_{iT}$</td>
<td>Potential outcome if $i$ receives treatment (i.e., in treatment group)</td>
</tr>
<tr>
<td>$Z$</td>
<td>Background factor(s) that affect $Y$ and may also affect $X$, and thus may serve as confounder(s)</td>
</tr>
<tr>
<td>$z$</td>
<td>Z-score</td>
</tr>
</tbody>
</table>
Acknowledgments

We are grateful for comments and suggestions received from Joe Bizup (Chapter 14), Taylor Boas (Part IV), Colin Elman (Chapter 13), Diana Kapiszewski (Chapter 13), Ryan Moore (Chapter 23), Max Palmer (Chapter 17), Laurel Smith-Doerr (Chapter 16), Peter Spiegler (Chapter 5), Arun Swamy (Chapter 5), and Susan Wishinsky (Chapter 11), as well as reviewers for Cambridge University Press. We are also grateful for our students, Christina Jarymowycz, Joshua Yesnowitz, Matthew Maguire, Sahar Abi-Hassan, and Cantay Caliskan, who worked on various aspects of the manuscript and related materials. Finally, we want to thank our editor at the Press, John Haslam, who shepherded this book along the path to publication, rendering wise counsel and keeping our noses to the proverbial grindstone.

Chapters 1–4, 6–8, 10, 12, and 17 draw on material published originally in Gerring’s Social Science Methodology: A Unified Framework (Cambridge University Press, 2012). Chapter 9 is based loosely on Gerring’s “Selecting Cases for Intensive Analysis: A Diversity of Goals and Methods” (Sociological Methods & Research, 2016, joint with Lee Cojocaru) and Case Study Research: Principles and Practices, 2nd edn. (Cambridge University Press, 2017). Readers may refer to these works for a more detailed treatment of these subjects.
Preface

Once upon a time, the practice of social science could be understood as the application of commonsense and intuition – something you might develop in the course of growing up. This is no longer true, or only partly true. Although commonsense and intuition are still useful, the social science disciplines have moved well beyond what can be understood without specialized training.

If you want to become an artist, musician, engineer – or pretty much anything, these days – developing your technique in these highly specialized areas is essential. It takes great dedication, countless hours of concentrated work, and professional guidance. The same may be said for social science in the contemporary era. One may mourn the death of the amateur social scientist. But one might as well reconcile oneself to the fact.

In response, methods courses have proliferated at both the undergraduate and graduate level. Likewise, methodological skills are in high demand in the social sciences and their cognate professions. Successful careers in government, communications, education, social work, business, law, and all of the policy fields require a solid grounding in methodology. Whether one is applying for graduate programs or for a job, the material covered in this book should stand one in good stead.

Indeed, a working knowledge of social science tools of analysis may prove more crucial for one’s career than whatever substantive knowledge one acquires in the course of a college education. What one knows is less important than what one can do, and what one can do depends on a working knowledge of methodology.

These developments may be viewed as part of a broader sea-change, driven by the rise of computers and the Internet. With sophisticated IT tools at our disposal, factual knowledge about a subject is no longer at a premium and can usually be obtained from a Google search or from a specially designed database in milliseconds. Likewise, any repetitive procedure can be programmed as a set of algorithms on a computer. This means that the value of an education is no longer in the facts or established protocols you might learn. This sort of knowledge can be produced by machines in a more timely and accurate fashion than by the human brain. Our value-added, as humans, stems from our capacity to identify important questions and think through practical solutions to those questions in a creative fashion. This is the function of a broadly pitched course on methodology and it is what this text is designed to convey.

The present text is appropriate for use in introductory or intermediate methods courses at the undergraduate, master’s, or doctoral level. It is designed to assist those who are attempting to make sense of social science as well as those who are...
Preface

conducting original research. We assume no prior methodological knowledge, though we do presume that the reader has some background in at least one field of social science, e.g., anthropology, communications, criminal justice, economics (including business, finance, and management), education, environmental policy, international development, law, political science, psychology, public health, public policy, social work, sociology, or urban planning.¹

We try to address key points of social science methodology in an applied fashion – so that readers can put these methods to work. Note that insofar as we can impact the societies we live in (in a conscious fashion) social science is indispensable. We can’t enhance economic growth, health, and education – or reduce poverty, crime, conflict, inequality, and global warming – without consulting the work of social scientists. To understand that work, and to conduct original research on these topics, an understanding of the methodological principles underlying this set of practices is indispensable. We hope that you will approach social science methodology not simply as a means for self-advancement (though there is surely nothing wrong with that!) but also as a set of tools for changing – and preserving – the world.

A Wide-Ranging Approach

In many textbook markets the offerings are fairly similar. A standard format has been developed over the years that everyone adheres to (more or less), and the courses that utilize these texts bear a strong resemblance to each other. There is scholarly consensus in the field about how to teach a subject.

This does not describe the topic at hand. Gazing out across the social science disciplines one finds a wide range of methodological approaches, reflected in a wide range of textbooks. As a service to the prospective reader (and instructor) it may be helpful to indicate how this volume differs from other textbooks in this crowded field – and why.

Some methods texts limit their purview to a specific discipline, e.g., political science, sociology, or economics. This may seem reasonable, and it allows one to focus on a set of substantive problems that orient a field. However, few substantive problems are confined to a single discipline. In order to learn about crime, for example, you will probably need to read across the fields of sociology, psychology, law, political science, economics, and criminology. The same is true for most other problems, which do not observe neat disciplinary boundaries.

Of course, important differences in theory and method characterize the disciplines. But it does not follow that one is well-served by a text that offers only one view of how to conduct social science. A narrow methodological training does not prepare one to integrate knowledge from other disciplines. To understand the range of literature on a topic and to think creatively about methods that might be applied to that topic it makes sense to adopt an ecumenical approach. Hence, this book focuses broadly on the methodological principles of social science rather than on methods practiced within a single discipline.
Preface

Some texts are focused primarily on quantitative methods, i.e., statistics or econometrics. While these are important skills, this approach has a tendency to reduce methodology to mathematics. And this, in turn, presents a narrow and technical vision of social science that is not faithful to the way in which social science is practiced (or, at any rate, to the way it should be practiced). Statistics are the handmaiden of methodology, not the other way around.

Some texts are focused exclusively on qualitative methods. This is a hard topic to define, and these books are varied in their content and approach. A few are strongly anti-positivist, meaning that they reject the scientific ideal as it has been understood in the natural sciences. While we agree with the standard critique of a narrowly positivist approach to social science we also think the natural sciences and social sciences share a good deal in common. In any case, a book that treats only qualitative components of social science is missing a good deal of the action. Both qualitative and quantitative approaches are required as part of everyone’s social science education. Certainly, they are both required in order to make sense of the social science literature on a subject.

One way to handle this problem is to include both qualitative and quantitative methods within a single text but to keep them separate, with the idea that the tools are distinct and each draws on a different epistemology (theory of knowledge). In our opinion, this claim is difficult to sustain: “qualitative” and “quantitative” tools tend to blend together and their epistemological traditions are not as far apart as they might seem. More important, a segregated approach to knowledge is not helpful to the advancement of social science. If knowledge on a topic is to grow it must be based on a unified epistemology that encompasses both qualitative and quantitative methods. This is the approach taken in the present text.

The most distinctive feature of this book is its wide-ranging approach to the subject. The text is intended to encompass all of the social science disciplines, qualitative and quantitative methods, descriptive and causal knowledge, and experimental and observational research designs. We also address the nuts and bolts of how to conduct research, as laid out below.

Naturally, there are some topics that we do not have time or space to engage. However, relative to most methods texts this one qualifies as highly inclusive, offering an entrée to myriad aspects of social science methodology. To our way of thinking, these topics are all essential. And they are also closely linked. While there are many ways to do good social science these diverse approaches also share certain common elements. Only by grasping the full extent of social science’s diversity can we glimpse its underlying unity.

Outline and Features

With a text of this size the reader may want to read strategically, focusing on chapters that are most relevant to your current work and interests, skipping or skimming chapters that cover topics about which you are already well-informed. A good textbook need not be read cover-to-cover.
However, readers should also be aware that the book is organized in a cumulative fashion, with later sections building on previous sections. Something may be lost if you peruse the text in a scattershot fashion.

Part I sets forth the basic building blocks of social science methodology. Chapter 1 introduces our topic, social science methodology, expanding on themes in the Preface and introducing several specific examples that will be referred to throughout the book. Subsequent chapters within this section focus on (2) arguments (including theories and hypotheses), (3) concepts and measures, and (4) analyses.

Part II focuses on causal arguments and analysis. This topic is broken down into chapters dealing with (5) causal frameworks, (6) causal hypotheses and analyses, (7) experimental research designs, (8) non-experimental research designs, (9) case study research designs, and (10) diverse tools of causal inference.

Part III deals with the process of research and the presentation of results. This includes (11) reading and reviewing the literature on a subject, (12) brainstorming (finding a research topic and a specific hypothesis), (13) data gathering, (14) writing, (15) public speaking, and (16) ethics.

Part IV deals with statistics. This is divided into several topics: (17) data management, (18) univariate statistics, (19) probability distributions, (20) statistical inference, (21) bivariate statistics, (22) regression, and (23) causal inference.

Every effort has been made to divide up these subjects in a way that makes logical sense and to avoid unnecessary redundancies. Of course, topics do not always neatly divide into separate chapters and sections. There is a holistic quality to social science methodology; diverse topics invariably bleed into one another. To assist the reader, we indicate where the reader might look for further elaboration of an issue. You may also consult the Detailed Table of Contents or the Index.

An objective of the book is to introduce readers to key terms of social science methodology. When a term is first introduced, or when it is formally defined, it is printed in bold. At the end of each chapter the reader will find a list of these bolded terms, which may be useful for purposes of review. In the Index, we indicate the page on which a term is defined by printing that number in bold.

The online materials for this book include series of questions and exercises for each chapter under the heading Inquiries. These inquiries serve a review function, summarizing the main points of the chapter. Some questions are speculative, building on the material presented but also moving beyond it. Instructors may draw on these inquiries to structure class discussion, to construct quizzes or exams, or for assignments.

In posing questions and constructing exercises we are sensitive to the fact that readers of the book have diverse disciplinary backgrounds. Consequently, many of the inquiries are presented in a manner that allows for tailoring the questions to the reader’s particular field of expertise. Rather than imposing a particular concept or theory on a methodological issue we might ask readers to choose a concept or theory with which they are familiar and employ it to address a question in their course of study.
Preface

An introductory textbook of modest length must deal with topics in an expeditious fashion. Accordingly, we have omitted many qualifications, caveats, and citations to the literature in favor of a streamlined approach. Although the treatment in this text is somewhat more detailed than that found in many textbooks it is still highly selective when placed within the context of scholarly work on these subjects. This is the cost of writing a short book on a long subject. Readers who choose to continue in some branch of social science should view this book as a point of departure on their methodological journey. The online materials include lists of suggested readings and web sites related to topics broached in each chapter, under the heading Resources. Consider these references as an invitation to further study.