Contents

List of Tables xiii

List of Figures xv

Preface xix

Support materials for the book xxi

1 Getting started 1

1.1 Conventions . . . 1

1.2 Introduction . . . 3

1.3 The Stata screen . . . 5

1.4 Using an existing dataset . . . 8

1.5 An example of a short Stata session . . . 10

1.6 Summary . . . 16

1.7 Exercises . . . 16

2 Entering data 19

2.1 Creating a dataset . . . 19

2.2 An example questionnaire . . . 21

2.3 Developing a coding system . . . 22

2.4 Entering data . . . 27

2.4.1 Labeling values . . . 30

2.5 Saving your dataset . . . 33

2.6 Checking the data . . . 35

2.7 Summary . . . 38

2.8 Exercises . . . 38

3 Preparing data for analysis 41

3.1 Introduction . . . 41

3.2 Planning your work . . . 41
3.3 *Creating value labels* . . . 47
3.4 Reverse-code variables . . . 48
3.5 Creating and modifying variables . . . 51
3.6 Creating scales . . . 56
3.7 Saving some of your data . . . 60
3.8 Summary . . . 62
3.9 Exercises . . . 63

4 Working with commands, do-files, and results 65

4.1 Introduction . . . 65
4.2 How Stata commands are constructed . . . 66
4.3 Getting the command from the menu system . . . 70
4.4 Saving your results . . . 77
4.5 Logging your command file . . . 78
4.6 Summary . . . 80
4.7 Exercises . . . 81

5 Descriptive statistics and graphs for one variable 83

5.1 Descriptive statistics and graphs . . . 83
5.2 Where is the center of a distribution? . . . 84
5.3 How dispersed is the distribution? . . . 88
5.4 Statistics and graphs—unordered categories . . . 90
5.5 Statistics and graphs—ordered categories and variables . . . 98
5.6 Statistics and graphs—quantitative variables . . . 100
5.7 Summary . . . 106
5.8 Exercises . . . 107

6 Statistics and graphs for two categorical variables 109

6.1 Relationship between categorical variables . . . 109
6.2 Cross-tabulation . . . 110
6.3 Chi-squared test . . . 112
6.3.1 Degrees of freedom . . . 114

6.3.2 Probability tables 115
6.4 Percentages and measures of association 117
6.5 Ordered categorical variables 119
6.6 Interactive tables 122
6.7 Tables—linking categorical and quantitative variables 124
6.8 Summary 127
6.9 Exercises 128
7 Tests for one or two means **131**
7.1 Introduction to tests for one or two means 131
7.2 Randomization 134
7.3 Random sampling 136
7.4 Hypotheses 136
7.5 One-sample test of a proportion 136
7.6 Two-sample test of a proportion 139
7.7 One-sample test of means 143
7.8 Two-sample test of group means 145
7.8.1 *Testing for unequal variances* 150
7.9 Repeated-measures t test 151
7.10 Power analysis 153
7.11 Nonparametric alternatives 156
7.11.1 Mann–Whitney two-sample rank-sum test 156
7.11.2 Nonparametric alternative: median test 157
7.12 Summary 158
7.13 Exercises 159
8 Bivariate regression and correlation **161**
8.1 Introduction to bivariate correlation and regression 161
8.2 Scattergrams 162
8.3 *Plotting the regression line* 167
8.4 Correlation 168
8.5 Regression 175

8.6 Spearman's rho: rank-order correlation for ordinal data 179
8.7 Summary 180
8.8 Exercises 181

9 Analysis of variance **183**

9.1 The logic of one-way analysis of variance 183
9.2 ANOVA example 184
9.3 ANOVA example using survey data 191
9.4 A nonparametric alternative to ANOVA 194
9.5 Analysis of covariance 197
9.6 Two-way ANOVA 204
9.7 Repeated-measures design 208
9.8 Intraclass correlation—measuring agreement 213
9.9 Summary 215
9.10 Exercises 216

10 Multiple regression **219**

10.1 Introduction to multiple regression 219
10.2 What is multiple regression? 220
10.3 The basic multiple regression command 220
10.4 Increment in R-squared: semipartial correlations 225
10.5 Is the dependent variable normally distributed? 226
10.6 Are the residuals normally distributed? 230
10.7 Regression diagnostic statistics 234
10.7.1 Outliers and influential cases 234
10.7.2 Influential observations: DFbeta 237
10.7.3 Combinations of variables may cause problems 238
10.8 Weighted data 240
10.9 Categorical predictors and hierarchical regression 242
10.10 Fundamentals of interaction 250
10.11 Summary 255
10.12 Exercises 257

11 Logistic regression **259**

11.1 Introduction to logistic regression 259

11.2 An example 260

11.3 What is an odds ratio and a logit? 264

11.3.1 The odds ratio 266

11.3.2 The logit transformation 266

11.4 Data used in rest of chapter 267

11.5 Logistic regression 268

11.6 Hypothesis testing 275

11.6.1 Testing individual coefficients 276

11.6.2 Testing sets of coefficients 277

11.7 Nested logistic regressions 279

11.8 Summary 281

11.9 Exercises 282

12 Measurement, reliability, and validity **283**

12.1 Overview of reliability and validity 283

12.2 Constructing a scale 284

12.2.1 Generating a mean score for each person 285

12.3 Reliability 286

12.3.1 Stability and test–retest reliability 289

12.3.2 Equivalence 290

12.3.3 Split-half and alpha reliability—internal consistency 290

12.3.4 Kuder–Richardson reliability for dichotomous items 293

12.3.5 Rater agreement—kappa (κ) 294

12.4 Validity 297

12.4.1 Expert judgment 297

12.4.2 Criterion-related validity 298

12.4.3 Construct validity 299

12.5 Factor analysis 300

12.6 PCF analysis . . . 305
12.6.1 Orthogonal rotation: varimax . . . 308
12.6.2 Oblique rotation: promax . . . 310
12.7 But we wanted one scale, not four scales . . . 311
12.7.1 Scoring our variable . . . 312
12.8 Summary . . . 313
12.9 Exercises . . . 314
13 Appendix: What's next? 315
13.1 Introduction to the appendix . . . 315
13.2 Resources . . . 315
13.2.1 Web resources . . . 316
13.2.2 Books about Stata . . . 317
13.2.3 Short courses . . . 318
13.2.4 Acquiring data . . . 319
13.3 Summary . . . 320
References 321
Author index 323
Subject index 325