

Contents

List of figures	xv
List of tables	xxiii
List of boxed tips	xxv
Preface	xxix
Acknowledgments	xxxiii
Support materials for the book	xxxv
1 Getting started	1
1.1 Conventions	1
1.2 Introduction	4
1.3 The Stata screen	7
1.4 Using an existing dataset	10
1.5 An example of a short Stata session	11
1.6 Video aids to learning Stata	18
1.7 Summary	19
1.8 Exercises	19
2 Entering data	21
2.1 Creating a dataset	21
2.2 An example questionnaire	24
2.3 Developing a coding system	25
2.4 Entering data using the Data Editor	29
2.4.1 Value labels	33
2.5 The Variables Manager	34
2.6 The Data Editor (Browse) view	40
2.7 Saving your dataset	41

2.8	Checking the data	43
2.9	Summary	50
2.10	Exercises	50
3	Preparing data for analysis	51
3.1	Introduction	51
3.2	Planning your work	52
3.3	Creating value labels	57
3.4	Reverse-code variables	60
3.5	Creating and modifying variables	65
3.6	Creating scales	70
3.7	Saving some of your data	74
3.8	Summary	75
3.9	Exercises	76
4	Working with commands, do-files, and results	77
4.1	Introduction	77
4.2	How Stata commands are constructed	78
4.3	Creating a do-file	83
4.4	Copying your results to a word processor	88
4.5	Logging your command file	90
4.6	Summary	91
4.7	Exercises	92
5	Descriptive statistics and graphs for one variable	93
5.1	Descriptive statistics and graphs	93
5.2	Where is the center of a distribution?	94
5.3	How dispersed is the distribution?	98
5.4	Statistics and graphs—unordered categories	100
5.5	Statistics and graphs—ordered categories and variables	110
5.6	Statistics and graphs—quantitative variables	112
5.7	Summary	120
5.8	Exercises	120

6	Statistics and graphs for two categorical variables	123
6.1	Relationship between categorical variables	123
6.2	Cross-tabulation	124
6.3	Chi-squared test	127
6.3.1	Degrees of freedom	129
6.3.2	Probability tables	129
6.4	Percentages and measures of association	132
6.5	Odds ratios when dependent variable has two categories	135
6.6	Ordered categorical variables	137
6.7	Interactive tables	140
6.8	Tables—linking categorical and quantitative variables	142
6.9	Power analysis when using a chi-squared test of significance	145
6.10	Summary	148
6.11	Exercises	148
7	Tests for one or two means	151
7.1	Introduction to tests for one or two means	151
7.2	Randomization	154
7.3	Random sampling	156
7.4	Hypotheses	156
7.5	One-sample test of a proportion	158
7.6	Two-sample test of a proportion	161
7.7	One-sample test of means	165
7.8	Two-sample test of group means	167
7.8.1	Testing for unequal variances	176
7.9	Repeated-measures t test	177
7.10	Power analysis	179
7.11	Nonparametric alternatives	187
7.11.1	Mann–Whitney two-sample rank-sum test	187
7.11.2	Nonparametric alternative: Median test	188
7.12	Video tutorial related to this chapter	189

7.13	Summary	189
7.14	Exercises	190
8	Bivariate correlation and regression	193
8.1	Introduction to bivariate correlation and regression	193
8.2	Scattergrams	194
8.3	Plotting the regression line	200
8.4	An alternative to producing a scattergram, binscatter	201
8.5	Correlation	205
8.6	Regression	210
8.7	Spearman's rho: Rank-order correlation for ordinal data	215
8.8	Power analysis with correlation	216
8.9	Summary	218
8.10	Exercises	218
9	Analysis of variance	221
9.1	The logic of one-way analysis of variance	221
9.2	ANOVA example	222
9.3	ANOVA example with nonexperimental data	231
9.4	Power analysis for one-way ANOVA	234
9.5	A nonparametric alternative to ANOVA	236
9.6	Analysis of covariance	238
9.7	Two-way ANOVA	249
9.8	Repeated-measures design	255
9.9	Intraclass correlation—measuring agreement	260
9.10	Power analysis with ANOVA	262
9.10.1	Power analysis for one-way ANOVA	263
9.10.2	Power analysis for two-way ANOVA	265
9.10.3	Power analysis for repeated-measures ANOVA	267
9.10.4	Summary of power analysis for ANOVA	269
9.11	Summary	270
9.12	Exercises	270

10	Multiple regression	273
10.1	Introduction to multiple regression	273
10.2	What is multiple regression?	274
10.3	The basic multiple regression command	275
10.4	Increment in R-squared: Semipartial correlations	279
10.5	Is the dependent variable normally distributed?	281
10.6	Are the residuals normally distributed?	284
10.7	Regression diagnostic statistics	290
10.7.1	Outliers and influential cases	290
10.7.2	Influential observations: DFbeta	292
10.7.3	Combinations of variables may cause problems	294
10.8	Weighted data	296
10.9	Categorical predictors and hierarchical regression	298
10.10	A shortcut for working with a categorical variable	307
10.11	Fundamentals of interaction	308
10.12	Nonlinear relations	315
10.12.1	Fitting a quadratic model	317
10.12.2	Centering when using a quadratic term	323
10.12.3	Do we need to add a quadratic component?	325
10.13	Power analysis in multiple regression	327
10.14	Summary	332
10.15	Exercises	333
11	Logistic regression	337
11.1	Introduction to logistic regression	337
11.2	An example	338
11.3	What is an odds ratio and a logit?	342
11.3.1	The odds ratio	344
11.3.2	The logit transformation	344
11.4	Data used in the rest of the chapter	345
11.5	Logistic regression	347

11.6	Hypothesis testing	357
11.6.1	Testing individual coefficients	358
11.6.2	Testing sets of coefficients	359
11.7	Margins: More on interpreting results from logistic regression	361
11.8	Nested logistic regressions	369
11.9	Power analysis when doing logistic regression	371
11.10	Next steps for using logistic regression and its extensions	374
11.11	Summary	374
11.12	Exercises	375
12	Measurement, reliability, and validity	377
12.1	Overview of reliability and validity	377
12.2	Constructing a scale	378
12.2.1	Generating a mean score for each person	379
12.3	Reliability	381
12.3.1	Stability and test–retest reliability	382
12.3.2	Equivalence	383
12.3.3	Split-half and alpha reliability—internal consistency	383
12.3.4	Kuder–Richardson reliability for dichotomous items	386
12.3.5	Rater agreement—kappa (κ)	387
12.4	Validity	389
12.4.1	Expert judgment	390
12.4.2	Criterion-related validity	391
12.4.3	Construct validity	391
12.5	Factor analysis	396
12.6	PCF analysis	400
12.6.1	Orthogonal rotation: Varimax	404
12.6.2	Oblique rotation: Promax	406
12.7	But we wanted one scale, not four scales	407
12.7.1	Scoring our variable	408

12.8	Summary	409
12.9	Exercises	410
13	Structural equation and generalized structural equation modeling	411
13.1	Linear regression using sem	412
13.1.1	Using the sem command directly	413
13.1.2	SEM and working with missing values	414
13.1.3	Exploring missing values and auxiliary variables	419
13.1.4	Getting auxiliary variables into your SEM command	421
13.2	A quick way to draw a regression model	421
13.3	The gsem command for logistic regression	425
13.3.1	Fitting the model using the logit command	425
13.3.2	Fitting the model using the gsem command	428
13.4	Path analysis and mediation	432
13.5	Conclusions and what is next for the sem command	437
13.6	Exercises	439
14	Working with missing values—multiple imputation	441
14.1	Working with missing values—multiple imputation	441
14.2	What variables do we include when doing imputations?	442
14.3	The nature of the problem	444
14.4	Multiple imputation and its assumptions about the mechanism for missingness	445
14.5	Multiple imputation	447
14.6	A detailed example	448
14.6.1	Preliminary analysis	449
14.6.2	Setup and multiple-imputation stage	452
14.6.3	The analysis stage	454
14.6.4	For those who want an R^2 and standardized β s	456
14.6.5	When impossible values are imputed	458
14.7	Summary	460
14.8	Exercises	461

15	An introduction to multilevel analysis	463
15.1	Questions and data for groups of individuals	463
15.2	Questions and data for a longitudinal multilevel application	464
15.3	Fixed-effects regression models	465
15.4	Random-effects regression models	466
15.5	An applied example	468
15.5.1	Research questions	468
15.5.2	Reshaping data to do multilevel analysis	469
15.6	A quick visualization of our data	472
15.7	Random-intercept model	473
15.7.1	Random intercept—linear model	473
15.7.2	Random-intercept model—quadratic term	476
15.7.3	Treating time as a categorical variable	480
15.8	Random-coefficients model	483
15.9	Including a time-invariant covariate	486
15.10	Summary	491
15.11	Exercises	492
16	Item response theory (IRT)	493
16.1	How are IRT measures of variables different from summated scales? .	494
16.2	Overview of three IRT models for dichotomous items	496
16.2.1	The one-parameter logistic (1PL) model	496
16.2.2	The two-parameter logistic (2PL) model	498
16.2.3	The three-parameter logistic (3PL) model	499
16.3	Fitting the 1PL model using Stata	500
16.3.1	The estimation	502
16.3.2	How important is each of the items?	504
16.3.3	An overall evaluation of our scale	506
16.3.4	Estimating the latent score	507
16.4	Fitting a 2PL IRT model	508
16.4.1	Fitting the 2PL model	509

16.5	The graded response model—IRT for Likert-type items	515
16.5.1	The data	515
16.5.2	Fitting our graded response model	517
16.5.3	Estimating a person's score	522
16.6	Reliability of the fitted IRT model	522
16.7	Using the Stata menu system	525
16.8	Extensions of IRT	528
16.9	Exercises	529
A	What's next?	531
A.1	Introduction to the appendix	531
A.2	Resources	531
A.2.1	Web resources	532
A.2.2	Books about Stata	534
A.2.3	Short courses	536
A.2.4	Acquiring data	537
A.2.5	Learning from the postestimation methods	538
A.3	Summary	539
	Glossary of acronyms	541
	Glossary of mathematical and statistical symbols	543
	References	545
	Author index	551
	Subject index	553