

# Contents

<b>Preface</b>	<b>xix</b>
<b>Support materials for the book</b>	<b>xxi</b>
<b>1 Getting started</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 The Stata screen . . . . .	3
1.3 Using an existing dataset . . . . .	7
1.4 An example of a short Stata session . . . . .	9
1.5 Conventions . . . . .	16
1.6 Chapter summary . . . . .	17
1.7 Exercises . . . . .	18
<b>2 Entering data</b>	<b>21</b>
2.1 Creating a dataset . . . . .	21
2.2 An example questionnaire . . . . .	23
2.3 Develop a coding system . . . . .	24
2.4 Entering data . . . . .	28
2.4.1 Labeling values . . . . .	30
2.5 Saving your dataset . . . . .	34
2.6 Checking the data . . . . .	35
2.7 Chapter summary . . . . .	38
2.8 Exercises . . . . .	38
<b>3 Preparing data for analysis</b>	<b>39</b>
3.1 Introduction . . . . .	39
3.2 Plan your work . . . . .	39

3.3	Create value labels . . . . .	44
3.4	Reverse-code variables . . . . .	45
3.5	Create and modify variables . . . . .	48
3.6	Create scales . . . . .	53
3.7	Save some of your data . . . . .	56
3.8	Summary . . . . .	58
3.9	Exercises . . . . .	59
<b>4</b>	<b>Working with commands, do-files, and results</b>	<b>61</b>
4.1	Introduction . . . . .	61
4.2	How Stata commands are constructed . . . . .	62
4.3	Getting the command from the menu system . . . . .	66
4.4	Saving your results . . . . .	72
4.5	Logging your command file . . . . .	73
4.6	Summary . . . . .	74
4.7	Exercises . . . . .	75
<b>5</b>	<b>Descriptive statistics and graphs for a single variable</b>	<b>77</b>
5.1	Descriptive statistics and graphs . . . . .	77
5.2	Where is the center of a distribution? . . . . .	78
5.3	How dispersed is the distribution? . . . . .	81
5.4	Statistics and graphs -- unordered categories . . . . .	82
5.5	Statistics and graphs -- ordered categories and variables . . . . .	88
5.6	Statistics and graphs -- quantitative variables . . . . .	90
5.7	Summary . . . . .	97
5.8	Exercises . . . . .	98
<b>6</b>	<b>Statistics and graphs for two categorical variables</b>	<b>101</b>
6.1	Relationship between categorical variables . . . . .	101
6.2	Cross-tabulation . . . . .	102
6.3	Chi-squared . . . . .	104
6.3.1	Degrees of freedom-- optional . . . . .	106

6.3.2	Probability tables—optional . . . . .	106
6.4	Percentages and measures of association . . . . .	109
6.5	Ordered categorical variables . . . . .	110
6.6	Interactive tables . . . . .	113
6.7	Tables—linking categorical and quantitative variables . . . . .	114
6.8	Summary . . . . .	117
6.9	Exercises . . . . .	118
<b>7</b>	<b>Tests for one or two means</b>	<b>119</b>
7.1	Tests for one or two means . . . . .	119
7.2	Randomization . . . . .	121
7.3	Hypotheses . . . . .	124
7.4	One-sample test of a proportion . . . . .	124
7.5	Two-sample test of a proportion . . . . .	127
7.6	One-sample test of means . . . . .	131
7.7	Two-sample test of group means . . . . .	133
7.7.1	Testing for unequal variances . . . . .	138
7.8	Repeated-measures t test . . . . .	139
7.9	Power analysis . . . . .	141
7.10	Nonparametric alternatives . . . . .	144
7.10.1	Mann-Whitney two-sample rank-sum test . . . . .	144
7.10.2	Nonparametric alternative: median test . . . . .	145
7.11	Summary . . . . .	145
7.12	Exercises . . . . .	146
<b>8</b>	<b>Bivariate correlation and regression</b>	<b>149</b>
8.1	Introduction to bivariate correlation and regression . . . . .	149
8.2	Scattergrams . . . . .	150
8.3	Plotting the regression line . . . . .	155
8.4	Correlation . . . . .	157
8.5	Regression . . . . .	163

8.6	Spearman's rho: rank-order correlation for ordinal data . . . . .	167
8.7	Alpha reliability . . . . .	168
8.8	Kappa as a measure of agreement for categorical data . . . . .	173
8.9	Summary . . . . .	174
8.10	Exercises . . . . .	175
<b>9</b>	<b>Analysis of variance (ANOVA)</b>	<b>177</b>
9.1	The logic of one-way analysis of variance . . . . .	177
9.2	ANOVA example . . . . .	178
9.3	ANOVA example using survey data . . . . .	185
9.4	A nonparametric alternative to ANOVA . . . . .	188
9.5	Analysis of covariance . . . . .	191
9.6	Two-way ANOVA . . . . .	198
9.7	Repeated-measures design . . . . .	202
9.8	Intraclass correlation—measuring agreement . . . . .	207
9.9	Summary . . . . .	208
9.10	Exercises . . . . .	209
<b>10</b>	<b>Multiple regression</b>	<b>211</b>
10.1	Introduction . . . . .	211
10.2	What is multiple regression? . . . . .	211
10.3	The basic multiple regression command . . . . .	212
10.4	Increment in R-squared: semipartial correlations . . . . .	216
10.5	Is the dependent variable normally distributed? . . . . .	218
10.6	Are the residuals normally distributed? . . . . .	220
10.7	Regression diagnostic statistics . . . . .	224
	10.7.1 Outliers and influential cases . . . . .	224
	10.7.2 Influential observations: dfbeta . . . . .	226
	10.7.3 Combinations of variables may cause problems . . . . .	227
10.8	Weighted data . . . . .	229
10.9	Categorical predictors and hierarchical regression . . . . .	232

10.10	Fundamentals of interaction . . . . .	240
10.11	Summary . . . . .	245
10.12	Exercises . . . . .	246
<b>11</b>	<b>Logistic regression</b>	<b>249</b>
11.1	Introduction . . . . .	249
11.2	An example . . . . .	250
11.3	What are an odds ratio and a logit? . . . . .	254
11.3.1	The odds ratio . . . . .	256
11.3.2	The logit transformation . . . . .	256
11.4	Data used in rest of chapter . . . . .	257
11.5	Logistic regression . . . . .	258
11.6	Hypothesis testing . . . . .	264
11.6.1	Testing individual coefficients . . . . .	264
11.6.2	Testing sets of coefficients . . . . .	265
11.7	Nested logistic regressions . . . . .	267
11.8	Summary . . . . .	268
11.9	Exercises . . . . .	270
<b>12</b>	<b>What's next?</b>	<b>271</b>
12.1	Introduction . . . . .	271
12.2	Resources . . . . .	271
12.2.1	Web resources . . . . .	272
12.2.2	Books on Stata . . . . .	273
12.2.3	Short courses . . . . .	274
12.2.4	Acquiring data . . . . .	275
12.3	Summary . . . . .	276
	<b>References</b>	<b>277</b>
	<b>Author index</b>	<b>279</b>
	<b>Subject index</b>	<b>281</b>