

Contents

List of Notes — VIII

- 1 Introduction — 1**
 - 1.1 Formatting — 1
 - 1.2 Graphic style — 2
 - 1.3 Version info — 3
 - 1.4 Online resources — 3
 - 1.5 Cheat sheet — 3

- 2 The first steps — 4**
 - 2.1 The graphical user interface (GUI) — 4
 - 2.2 Opening stata files — 5
 - 2.3 Importing non-Stata file formats — 7
 - 2.4 Entering data manually — 7
 - 2.5 Using preinstalled data — 8
 - 2.6 Saving and exporting data — 9
 - 2.7 The basic workflow — 10
 - 2.8 Do-files — 11
 - 2.9 Delimit and line breaks* — 14

- 3 Cleaning and preparing data — 16**
 - 3.1 Getting to know your data — 16
 - 3.2 Variable names and labels — 18
 - 3.3 Labeling values — 19
 - 3.4 IDs and unique identifiers — 21
 - 3.5 Missing values — 24
 - 3.6 Creating new variables — 26
 - 3.6.1 Special functions — 28
 - 3.7 The if qualifier — 30
 - 3.8 Changing and replacing variables — 32
 - 3.9 Removing observations and variables — 35
 - 3.10 Cleaning data systematically — 36
 - 3.11 Combining datasets* — 38
 - 3.11.1 Appending datasets — 38
 - 3.11.2 One-to-One Merge — 39
 - 3.11.3 Many-to-One Merge — 41
 - 3.11.4 One-to-Many Merge — 42
 - 3.11.5 All pairwise combinations — 42
 - 3.12 Reshaping data* — 43

4	Describing data — 46
4.1	Summarizing information — 47
4.2	Using stored results* — 49
4.3	Histograms — 51
4.4	Boxplots — 53
4.5	Simple bar charts — 55
4.6	Scatterplots — 56
4.7	Frequency tables — 58
4.8	Summarizing information by categories — 61
4.9	Editing and exporting graphs — 64
4.9.1	Combining graphs — 65
4.10	Correlations — 67
4.11	Testing for normality — 68
4.12	t-test for groups* — 69
4.13	Weighting* — 71
5	Introduction to causal analysis — 73
5.1	Correlation and causation — 73
5.2	Causal graphs — 74
5.3	Estimating causal effects — 77
5.4	What does “controlling” actually mean?* — 79
6	Regression analysis — 83
6.1	Research question — 83
6.2	What is a regression? — 84
6.3	Binary independent variable — 85
6.4	Ordinal independent variable — 87
6.5	Metric independent variable — 91
6.6	Interaction effects* — 94
6.6.1	The classic way — 96
6.6.2	Marginal effects — 96
6.6.3	Predicted values — 97
6.6.4	Separate analyses by subgroups — 99
6.7	Standardized regression coefficients* — 100
7	Regression diagnostics — 102
7.1	Exogeneity — 102
7.2	Random sampling — 103
7.3	Linearity in parameters — 103
7.3.1	Solutions — 106
7.4	Multicollinearity — 107
7.4.1	Solutions — 108

7.5	Heteroscedasticity — 108
7.5.1	Solutions — 110
7.6	Influential observations — 113
7.6.1	Dfbetas — 114
7.6.2	Cook's distance — 115
7.7	Summary — 117
8	Logistic regression* — 119
8.1	Introduction — 119
8.2	Control variables — 125
8.3	Nested Models — 129
8.4	Diagnostics — 130
8.4.1	Model misspecification — 130
8.4.2	Sample size and empty cells — 130
8.4.3	Multicollinearity — 131
8.4.4	Influential observations — 132
9	Matching — 134
9.1	Simulating an experiment — 134
9.2	Propensity score matching — 135
9.3	Matching diagnostics — 138
9.3.1	Common support — 138
9.3.2	Balancing of covariates — 139
10	Reporting results — 141
10.1	Tables — 141
10.2	Graphs — 144
11	Writing a seminar paper — 150
11.1	The basic structure — 150
11.2	Master do-files — 152
12	The next steps — 153
12.1	Online sources and manuals — 153
12.2	Books — 153
References — 155	
Copyright — 157	
Index — 158	