

Contents

	Acknowledgements	v
	List of tables	xiii
	List of figures	xv
	Preface	xvii
1	Introduction	1
1.1	Using this book	2
1.2	Overview of this book	3
1.3	Listing observations in this book	4
2	Reading and writing datasets	9
2.1	Introduction	10
2.2	Reading Stata datasets	14
2.3	Saving Stata datasets	16
2.4	Reading comma-separated and tab-separated files	18
2.5	Reading space-separated files	20
2.6	Reading fixed-column files	22
2.7	Reading fixed-column files with multiple lines of raw data per observation	26
2.8	Reading SAS XPORT files	29
2.9	Common errors reading files	30
2.10	Entering data directly into the Stata Data Editor	33
2.11	Saving comma-separated and tab-separated files	40
2.12	Saving space-separated files	41
2.13	Saving SAS XPORT files	43
3	Data cleaning	45
3.1	Introduction	46

3.2	Double data entry	47
3.3	Checking individual variables	50
3.4	Checking categorical by categorical variables	54
3.5	Checking categorical by continuous variables	56
3.6	Checking continuous by continuous variables	60
3.7	Correcting errors in data	63
3.8	Identifying duplicates	67
3.9	Final thoughts on data cleaning	75
4	Labeling datasets	77
4.1	Introduction	78
4.2	Describing datasets	78
4.3	Labeling variables	84
4.4	Labeling values	86
4.5	Labeling utilities	92
4.6	Labeling variables and values in different languages	97
4.7	Adding comments to your dataset using notes	102
4.8	Formatting the display of variables	106
4.9	Changing the order of variables in a dataset	110
5	Creating variables	115
5.1	Introduction	116
5.2	Creating and changing variables	116
5.3	Numeric expressions and functions	120
5.4	String expressions and functions	121
5.5	Recoding	125
5.6	Coding missing values	130
5.7	Dummy variables	133
5.8	Date variables	137
5.9	Date-and-time variables	144
5.10	Computations across variables	150
5.11	Computations across observations	152

5.12	More examples using the egen command	155
5.13	Converting string variables to numeric variables	157
5.14	Converting numeric variables to string variables	163
5.15	Renaming and ordering variables	166
6	Combining datasets	173
6.1	Introduction	174
6.2	Appending: Appending datasets	174
6.3	Appending: Problems	178
6.4	Merging: One-to-one match-merging	189
6.5	Merging: One-to-many match-merging	195
6.6	Merging: Merging multiple datasets	199
6.7	Merging: Update merges	203
6.8	Merging: Additional options when merging datasets	206
6.9	Merging: Problems merging datasets	211
6.10	Joining datasets	216
6.11	Crossing datasets	218
7	Processing observations across subgroups	221
7.1	Introduction	222
7.2	Obtaining separate results for subgroups	222
7.3	Computing values separately by subgroups	224
7.4	Computing values within subgroups: Subscripting observations	228
7.5	Computing values within subgroups: Computations across observations	234
7.6	Computing values within subgroups: Running sums	236
7.7	Computing values within subgroups: More examples	238
7.8	Comparing the by and tsset commands	244
8	Changing the shape of your data	247
8.1	Introduction	248
8.2	Wide and long datasets	248
8.3	Introduction to reshaping long to wide	258

8.4	Reshaping long to wide: Problems	261
8.5	Introduction to reshaping wide to long	262
8.6	Reshaping wide to long: Problems	266
8.7	Multilevel datasets	271
8.8	Collapsing datasets	274
9	Programming for data management	277
9.1	Introduction	278
9.2	Tips on long-term goals in data management	279
9.3	Executing do-files and making log files	282
9.4	Automating data checking	289
9.5	Combining do-files	292
9.6	Introducing Stata macros	296
9.7	Manipulating Stata macros	300
9.8	Repeating commands by looping over variables	303
9.9	Repeating commands by looping over numbers	310
9.10	Repeating commands by looping over anything	312
9.11	Accessing results saved from Stata commands	314
9.12	Saving results of estimation commands as data	318
9.13	Writing Stata programs	323
10	Additional resources	329
10.1	Online resources for this book	330
10.2	Finding and installing additional programs	330
10.3	More online resources	339
A	Common elements	341
A.1	Introduction	342
A.2	Overview of Stata syntax	342
A.3	Working across groups of observations with by	344
A.4	Comments	346
A.5	Data types	347
A.6	Logical expressions	357

A.7 Functions 361

A.8 Subsetting observations with if and in 364

A.9 Subsetting observations and variables with keep and drop 367

A.10 Missing values 370

A.11 Referring to variable lists 374

Subject index **379**

21 *Meaning of newvar depending on the value inserted for X* 231

22 *Values assigned to newvar based on the value inserted for X* 233

23 *Expressions to replace X and the meaning that newvar would have* 234

24 *Expressions to replace X and the meaning that newvar would have* 237