

Contents of User's Guide

BASICS

1	Read this — it will help	3
1.1	Getting Started with Stata	4
1.2	The User's Guide and the Reference manuals	4
1.2.1	Cross-referencing	5
1.2.2	The index	5
1.2.3	The subject table of contents	5
1.2.4	Typography	5
1.3	What's new	6
1.3.1	What's big	7
1.3.2	What's useful	9
1.3.3	What's convenient	10
1.3.4	What was needed	10
1.3.5	What's faster	10
1.3.6	What's new in time-series analysis	11
1.3.7	What's new in cross-sectional time-series analysis	11
1.3.8	What's new in survival analysis	12
1.3.9	What's new in survey analysis	13
1.3.10	What's new in cluster analysis	14
1.3.11	What's new in statistics useful in all fields	14
1.3.12	What's new in data management	19
1.3.13	What's new in expressions and functions	22
1.3.14	What's new in display formats	25
1.3.15	What's new in programming	25
1.3.16	What's new in the user interface	27
1.3.17	What's more	28
2	Resources for learning and using Stata	29
2.1	Overview	29
2.2	The http://www.stata.com web site	30
2.3	The http://www.stata-press.com web site	30
2.4	The Stata listserver	31
2.5	The Stata Journal and the Stata Technical Bulletin	31
2.6	Updating and adding features from the web	32
2.6.1	Official updates	32
2.6.2	Unofficial updates	32
2.7	NetCourses	33
2.8	Books and other support materials	33
2.9	Technical support	34
2.9.1	Register your software	34
2.9.2	Before contacting technical support	34
2.9.3	Technical support by email	34
2.9.4	Technical support by phone or fax	35
2.9.5	Comments and suggestions for our technical staff	35
3	A brief description of Stata	37

4	Flavors of Stata	41
4.1	Platforms	41
4.2	Stata/SE, Intercooled Stata, and Small Stata	41
4.2.1	Determining which version you own	42
4.2.2	Determining which version is installed	42
4.3	Size limits comparison of Stata/SE, Intercooled Stata, and Small Stata	42
4.4	Speed comparison of Stata/SE, Intercooled Stata, and Small Stata	43
4.5	Features of Stata/SE	43
5	Starting and stopping Stata	45
5.1	Starting Stata	45
5.2	Verifying that Stata is correctly installed	46
5.3	Exiting Stata	47
5.4	Features worth learning about	47
5.4.1	Windows	47
5.4.2	Macintosh	48
5.4.3	Unix	48
6	Troubleshooting starting and stopping Stata	51
6.1	If Stata does not start	51
6.2	verinst problems	52
6.3	Troubleshooting Windows	52
6.4	Troubleshooting Macintosh	52
6.5	Troubleshooting Unix	52
7	Setting the size of memory	53
7.1	Memory size considerations	53
7.2	Setting memory size on the fly: Stata/SE	53
7.2.1	Advice on setting maxvar	54
7.2.2	Advice on setting matsize	55
7.2.3	Advice on setting memory	56
7.3	Setting memory size on the fly: Intercooled Stata	57
7.4	The memory command	57
7.5	Virtual memory and speed considerations	59
7.6	An issue when returning memory to Unix	59
8	Stata's online help and search facilities	61
8.1	Introduction	61
8.2	help: Stata's online manual pages	61
8.3	search: Stata's online index	63
8.4	Accessing help and search from the Help menu	65
8.5	More on search	65
8.6	More on help	66
8.7	help contents: Stata's online table of contents	67
8.8	search: All the details	67
8.8.1	How search works	67
8.8.2	Author searches	68
8.8.3	Entry id searches	68
8.8.4	FAQ searches	69
8.8.5	Return codes	69
8.9	net search: Searching net resources	70

9	Stata's sample datasets	71
10	–more– conditions	73
10.1	Description	73
10.2	set more off	73
10.3	The more programming command	73
11	Error messages and return codes	75
11.1	Making mistakes	75
11.1.1	Mistakes are forgiven	75
11.1.2	Mistakes stop user-written programs and do-files	75
11.1.3	Advanced programming to tolerate errors	76
11.2	The return message for obtaining command timings	76
12	The Break key	79
12.1	Making Stata stop what it is doing	79
12.2	Side-effects of pressing Break	80
12.3	Programming considerations	80
13	Keyboard use	81
13.1	Description	81
13.2	F-keys	81
13.3	Editing keys in Stata for Windows, Macintosh, and Unix(GUI)	83
13.4	Editing keys in Stata for Unix(console)	84
13.5	Editing previous lines in Stata for all operating systems	85
13.6	Tab expansion of variable names	86

ELEMENTS

14	Language syntax	89
14.1	Overview	89
14.1.1	varlist	89
14.1.2	by varlist:	90
14.1.3	if exp	92
14.1.4	in range	93
14.1.5	=exp	94
14.1.6	weight	95
14.1.7	options	96
14.1.8	numlist	98
14.2	Abbreviation rules	99
14.2.1	Command abbreviation	99
14.2.2	Option abbreviation	100
14.2.3	Variable-name abbreviation	100
14.3	Naming conventions	101
14.4	varlists	101
14.4.1	Lists of existing variables	101
14.4.2	Lists of new variables	102
14.4.3	Time-series varlists	104
14.5	by varlist: construct	106

14.6	File-naming conventions	109
14.6.1	A special note for Macintosh users	110
14.6.2	A special note for Unix users	111
15	Data	113
15.1	Data and datasets	113
15.2	Numbers	114
15.2.1	Missing values	114
15.2.2	Numeric storage types	118
15.3	Dates	118
15.4	Strings	119
15.4.1	Strings containing identifying data	119
15.4.2	Strings containing categorical data	119
15.4.3	Strings containing numeric data	120
15.4.4	String storage types	120
15.5	Formats: controlling how data are displayed	120
15.5.1	Numeric formats	120
15.5.2	European numeric formats	123
15.5.3	Date formats	124
15.5.4	Time-series formats	125
15.5.5	String formats	126
15.6	Dataset, variable, and value labels	127
15.6.1	Dataset labels	127
15.6.2	Variable labels	128
15.6.3	Value labels	129
15.7	Notes attached to data	134
15.8	Characteristics	135
16	Functions and expressions	137
16.1	Overview	137
16.2	Operators	138
16.2.1	Arithmetic operators	138
16.2.2	String operators	138
16.2.3	Relational operators	139
16.2.4	Logical operators	140
16.2.5	Order of evaluation, all operators	141
16.3	Functions	141
16.4	System variables (_variables)	142
16.5	Accessing coefficients and standard errors	142
16.5.1	Simple models	143
16.5.2	ANOVA and MANOVA models	143
16.5.3	Multiple-equation models	143
16.6	Accessing results from Stata commands	144
16.7	Explicit subscripting	145
16.7.1	Generating lags and leads	145
16.7.2	Subscripting within groups	146
16.8	Time-series operators	149
16.8.1	Generating lags and leads	149
16.8.2	Operators within groups	149
16.9	Label values	150
16.10	Precision and problems therein	150

17	Matrix expressions	153
17.1	Overview	153
17.1.1	Definition of a matrix	153
17.1.2	matsize	154
17.2	Row and column names	154
17.2.1	The purpose of row and column names	155
17.2.2	Three-part names	157
17.2.3	Setting row and column names	158
17.2.4	Obtaining row and column names	159
17.3	Vectors and scalars	159
17.4	Inputting matrices by hand	159
17.5	Accessing matrices created by Stata commands	160
17.6	Creating matrices by accumulating data	161
17.7	Matrix operators	161
17.8	Matrix functions	162
17.9	Subscripting	163
17.10	Using matrices in scalar expressions	164
18	Printing and preserving output	165
18.1	Overview	165
18.1.1	Starting and closing logs	166
18.1.2	Appending to an existing log	168
18.1.3	Temporarily suspending and resuming logging	168
18.2	Placing comments in logs	169
18.3	Logging only what you type	169
18.4	The log-button alternative	170
18.5	Printing logs	170
19	Do-files	171
19.1	Description	171
19.1.1	Version	172
19.1.2	Comments and blank lines in do-files	173
19.1.3	Long lines in do-files	174
19.1.4	Error handling in do-files	175
19.1.5	Logging the output of do-files	177
19.1.6	Preventing –more– conditions	177
19.2	Calling other do-files	178
19.3	Ways to run a do-file (Stata for Windows)	178
19.4	Ways to run a do-file (Stata for Macintosh)	179
19.5	Ways to run a do-file (Stata for Unix)	179
19.6	Programming with do-files	180
19.6.1	Argument passing	180
19.6.2	Suppressing output	181
20	Ado-files	183
20.1	Description	183
20.2	What is an ado-file?	183
20.3	How can I tell if a command is built-in or an ado-file?	184
20.4	Can I look at an ado-file?	184

20.5	Where does Stata look for ado-files?	185
20.5.1	Where are the official ado directories?	185
20.5.2	Where is my personal ado directory?	185
20.6	How do I install an addition?	186
20.7	How do I add my own ado-files?	187
20.8	How do I install official updates?	187
21	Programming Stata	189
21.1	Description	190
21.2	Relationship between a program and a do-file	190
21.3	Macros	193
21.3.1	Local macros	194
21.3.2	Global macros	194
21.3.3	The difference between local and global macros	195
21.3.4	Macros and expressions	196
21.3.5	Double quotes	198
21.3.6	Extended macro functions	199
21.3.7	Macro expansion operators	203
21.3.8	Advanced local macro manipulation	205
21.3.9	Advanced global macro manipulation	207
21.3.10	Constructing Windows filenames using macros	208
21.3.11	Accessing system values	208
21.3.12	Referencing characteristics	209
21.4	Program arguments	209
21.4.1	Renaming positional arguments	212
21.4.2	Incrementing through positional arguments	213
21.4.3	Using macro shift	214
21.4.4	Parsing standard Stata syntax	215
21.4.5	Parsing immediate commands	217
21.4.6	Parsing nonstandard syntax	217
21.5	Scalars and matrices	218
21.6	Temporarily destroying the data in memory	218
21.7	Temporary objects	219
21.7.1	Temporary variables	219
21.7.2	Temporary scalars and matrices	219
21.7.3	Temporary files	220
21.8	Accessing results calculated by other programs	220
21.9	Accessing results calculated by estimation commands	223
21.10	Saving results	224
21.10.1	Saving results in r()	225
21.10.2	Saving results in e()	226
21.10.3	Saving results in s()	228
21.11	Ado-files	229
21.11.1	Version	231
21.11.2	Comments and long lines in ado-files	231
21.11.3	Debugging ado-files	231
21.11.4	Local subroutines	232
21.11.5	Development of a sample ado-command	233
21.11.6	Writing online help	239
21.12	A compendium of useful commands for programmers	243
21.13	References	243

22	Immediate commands	245
22.1	Overview	245
22.1.1	Examples	246
22.1.2	A list of the immediate commands	247
22.2	The display command	248
23	Estimation and post-estimation commands	249
23.1	All estimation commands work the same way	249
23.2	Standard syntax	251
23.3	Replaying prior results	252
23.4	Cataloging estimation results	253
23.5	Specifying the estimation subsample	255
23.6	Specifying the width of confidence intervals	255
23.7	Obtaining the variance-covariance matrix	256
23.8	Obtaining predicted values	257
23.8.1	predict can be used after any estimation command	258
23.8.2	Making in-sample predictions	259
23.8.3	Making out-of-sample predictions	259
23.8.4	Obtaining standard errors, tests, and confidence intervals for predictions	260
23.9	Accessing estimated coefficients	261
23.10	Performing hypothesis tests on the coefficients	263
23.10.1	Linear tests	263
23.10.2	test can be used after any estimation command	264
23.10.3	Likelihood-ratio tests	266
23.10.4	Nonlinear Wald tests	267
23.11	Obtaining linear combinations of coefficients	267
23.12	Obtaining nonlinear combinations of coefficients	269
23.13	Obtaining marginal effects	269
23.14	Obtaining robust variance estimates	270
23.15	Obtaining scores	275
23.16	Weighted estimation	278
23.16.1	Frequency weights	278
23.16.2	Analytic weights	279
23.16.3	Sampling weights	280
23.16.4	Importance weights	282
23.17	A list of post-estimation commands	282
23.18	References	283

ADVICE

24	Commands to input data	287
24.1	Six ways to input data	287
24.2	Eight rules for determining which input method to use	287
24.2.1	If you wish to enter data interactively: Rule 1	288
24.2.2	If the dataset is in binary format: Rule 2	288
24.2.3	If the data are simple: Rule 3	289
24.2.4	If the dataset is formatted and the formatting is significant: Rule 4	290
24.2.5	If there are no string variables: Rule 5	291
24.2.6	If all the string variables are enclosed in quotes: Rule 6	291
24.2.7	If the unlimited strings have no blanks: Rule 7	292
24.2.8	If you make it to here: Rule 8	293

24.3	If you run out of memory	293
24.4	Transfer programs	293
24.5	ODBC sources	294
24.6	References	294
25	Commands for combining data	295
26	Commands for dealing with strings	297
26.1	Description	297
26.2	Categorical string variables	297
26.3	Mistaken string variables	298
26.4	Complex strings	299
27	Commands for dealing with dates	301
27.1	Overview	301
27.2	Dates	302
27.2.1	Inputting dates	302
27.2.2	Conversion into elapsed dates	303
27.2.2.1	The mdy() function	303
27.2.2.2	The date() function	304
27.2.3	Displaying dates	306
27.2.4	Other date functions	309
27.2.5	Specifying particular dates (date literals)	309
27.3	Time-series dates	310
27.3.1	Inputting time variables	311
27.3.2	Specifying particular dates (date literals)	312
27.3.3	Time-series formats	313
27.3.4	Translating between time units	313
27.3.5	Extracting components of time	314
27.3.6	Creating time variables	314
27.3.7	Setting the time variable	314
27.3.8	Selecting periods of time	314
27.3.9	The %tg format	315
28	Commands for dealing with categorical variables	317
28.1	Continuous, categorical, and indicator variables	317
28.1.1	Converting continuous to indicator variables	318
28.1.2	Converting continuous to categorical variables	318
28.1.3	Converting categorical to indicator variables	320
28.2	Using indicator variables in estimation	322
28.2.1	Testing the significance of indicator variables	323
28.2.2	Importance of omitting one of the indicators	324
29	Overview of Stata estimation commands	327
29.1	Introduction	327
29.2	Linear regression with simple error structures	328
29.3	ANOVA, ANCOVA, MANOVA, and MANCOVA	320
29.4	Generalized linear models	330
29.5	Binary outcome qualitative dependent variable models	330

29.6	Conditional logistic regression	332
29.7	Multiple outcome qualitative dependent variable models	332
29.8	Simple count dependent variable models	333
29.9	Linear regression with heteroskedastic errors	333
29.10	Stochastic frontier models	333
29.11	Linear regression with systems of equations (correlated errors)	334
29.12	Models with endogenous sample selection	334
29.13	Models with time-series data	335
29.14	Panel-data models	335
29.14.1	Linear regression with panel data	335
29.14.2	Censored linear regression with panel data	337
29.14.3	Generalized linear models with panel data	337
29.14.4	Qualitative dependent variable models with panel data	337
29.14.5	Count dependent variable models with panel data	338
29.14.6	Random-coefficient models with panel data	338
29.15	Survival-time (failure-time) models	338
29.16	Commands for estimation with survey data	338
29.17	Multivariate analysis	339
29.18	Pharmacokinetic data	340
29.19	Cluster analysis	340
29.20	Not elsewhere classified	341
29.21	Have we forgotten anything?	341
29.22	References	341
30	Overview of survey estimation	343
30.1	Introduction	343
30.2	Accounting for the sample design in survey analyses	345
30.2.1	Multistage sample designs	346
30.2.2	Finite population corrections	347
30.2.3	Design effects: deff and deft	348
30.3	Example of the effects of weights, clustering, and stratification	348
30.3.1	Halfway isn't enough: the importance of stratification and clustering ..	349
30.4	Linear regression and other models	351
30.5	Pseudo-likelihoods	352
30.6	Building your own survey estimator using ml	353
30.7	Hypothesis testing using test and testnl	353
30.8	Estimation of linear and nonlinear combinations of parameters	354
30.9	Two-way contingency tables	355
30.10	Differences between the svy commands and other commands	356
30.11	References	358
31	Commands everyone should know	360
31.1	Forty-one commands	360
31.2	The by construct	361
32	Using the Internet to keep up to date	362
32.1	Overview	363
32.2	Sharing datasets (and other files)	363

32.3	Official updates	363
32.3.1	Example	364
32.3.2	Updating ado-files	365
32.3.3	Frequently asked questions about updating the ado-files	366
32.3.4	Updating the executable	366
32.3.5	Frequently asked questions about updating the executable	367
32.3.6	Updating both ado-files and the executable	367
32.4	Downloading and managing additions by users	368
32.4.1	Downloading files	368
32.4.2	Managing files	370
32.4.3	Finding files to download	370
32.5	Making your own download site	371
 Author index		 373
 Subject index		 387