

Contents

Preface	xxiii
I General Information	1
1 Introduction	3
1.1 What is this book about?	3
1.2 Which models are considered?	4
1.3 Who is this book for?	5
1.4 How is the book organized?	5
1.5 What software do you need?	7
1.5.1 Updating Stata 8	7
1.5.2 Installing SPost	8
Installing SPost using net search	8
Installing SPost using net install	10
1.5.3 What if commands do not work?	10
1.5.4 Uninstalling SPost	11
1.5.5 Additional files available on the web site	11
1.6 Where can I learn more about the models?	12
2 Introduction to Stata	13
2.1 The Stata interface	14
Changing the scrollbar buffer size	16
Changing the display of variable names in the Variables window	16
2.2 Abbreviations	17
2.3 How to get help	17

2.3.1	Online help	17
2.3.2	Manuals	18
2.3.3	Other resources	19
2.4	The working directory	19
2.5	Stata file types	19
2.6	Saving output to log files	20
	Options	20
2.6.1	Closing a log file	21
2.6.2	Viewing a log file	21
2.6.3	Converting from SMCL to plain text or PostScript	21
2.7	Using and saving datasets	22
2.7.1	Data in Stata format	22
2.7.2	Data in other formats	23
2.7.3	Entering data by hand	23
2.8	Size limitations on datasets*	23
2.9	do-files	24
2.9.1	Adding comments	25
2.9.2	Long lines	26
2.9.3	Stopping a do-file while it is running	26
2.9.4	Creating do-files	26
	Using Stata's do-file editor	26
	Using other editors to create do-files	26
2.9.5	A recommended structure for do-files	27
2.10	Using Stata for serious data analysis	28
2.11	The syntax of Stata commands	30
2.11.1	Commands	31
2.11.2	Variable lists	31
2.11.3	if and in qualifiers	32
	Examples of if qualifier	33
2.11.4	Options	34

2.12	Managing data	34
2.12.1	Looking at your data	34
2.12.2	Getting information about variables	34
2.12.3	Missing values	37
2.12.4	Selecting observations	38
2.12.5	Selecting variables	38
2.13	Creating new variables	39
2.13.1	generate command	39
2.13.2	replace command	40
2.13.3	recode command	40
2.13.4	Common transformations for RHS variables	42
	Breaking a categorical variable into a set of binary variables . . .	42
	More examples of creating binary variables	44
	Nonlinear transformations	44
	Interaction terms	45
2.14	Labeling variables and values	45
2.14.1	Variable labels	46
2.14.2	Value labels	46
2.14.3	notes command	48
2.15	Global and local macros	48
2.16	Graphics	49
2.16.1	The graph command	50
2.16.2	Displaying previously drawn graphs	56
2.16.3	Printing graphs	56
2.16.4	Combining graphs	56
2.17	A brief tutorial	58
	A batch version	65
3	Estimation, Testing, Fit, and Interpretation	67
3.1	Estimation	68

3.1.1	Stata's output for ML estimation	68
3.1.2	ML and sample size	69
3.1.3	Problems in obtaining ML estimates	69
3.1.4	The syntax of estimation commands	70
	Variable lists	70
	Specifying the estimation sample	71
	Options	73
3.1.5	Reading the output	75
	Header	75
	Estimates and standard errors	76
3.1.6	Reformatting output with estimates table	76
3.1.7	Reformatting output with outreg	78
3.1.8	Alternative output with listcoef	79
	Options for types of coefficients	79
	Other options	80
	Standardized coefficients	80
	Factor and percent change	83
3.1.9	Storing estimation results	83
3.2	Post-estimation analysis	83
3.3	Testing	84
3.3.1	Wald tests	84
	The accumulate option	86
3.3.2	LR tests	86
	Avoiding invalid LR tests	87
3.4	Measures of fit	87
	Syntax of fitstat	88
	Options	88
	Models and measures	89
	Example of fitstat	89
	Methods and formulas for fitstat	90

3.5 Interpretation 95

3.5.1 Approaches to interpretation 98

3.5.2 Predictions using predict 98

3.5.3 Overview of prchange, prgen, prtab, and prvalue 99

Specifying the levels of variables 100

Options controlling output 101

3.5.4 Syntax for prchange 101

Options 102

3.5.5 Syntax for prgen 102

Options 102

Variables generated 103

3.5.6 Syntax for prtab 103

Options 103

3.5.7 Syntax for prvalue 104

Options 104

3.5.8 Computing marginal effects using mfx compute 104

3.6 Next steps 105

II Models for Specific Kinds of Outcomes 107

4 Models for Binary Outcomes 109

4.1 The statistical model 110

4.1.1 A latent variable model 110

4.1.2 A nonlinear probability model 113

4.2 Estimation using logit and probit 114

Variable lists 114

Specifying the estimation sample 114

Weights 114

Options 115

Example 115

4.2.1	Observations predicted perfectly	117
4.3	Hypothesis testing with test and lrtest	118
4.3.1	Testing individual coefficients	118
	One and two-tailed tests	119
	Testing single coefficients using test	119
	Testing single coefficients using lrtest	120
4.3.2	Testing multiple coefficients	121
	Testing multiple coefficients using test	121
	Testing multiple coefficients using lrtest	122
4.3.3	Comparing LR and Wald tests	122
4.4	Residuals and influence using predict	123
4.4.1	Residuals	124
	Example	125
4.4.2	Influential cases	128
4.5	Scalar measures of fit using fitstat	129
	Example	129
4.6	Interpretation using predicted values	131
4.6.1	Predicted probabilities with predict	131
4.6.2	Individual predicted probabilities with prvalue	134
4.6.3	Tables of predicted probabilities with prtab	136
4.6.4	Graphing predicted probabilities with prgen	137
4.6.5	Changes in predicted probabilities	139
	Marginal change	139
	Discrete change	141
4.7	Interpretation using odds ratios with listcoef	145
	Multiplicative coefficients	147
	Effect of the base probability	147
	Percent change in the odds	148
4.8	Other commands for binary outcomes	149

5	Models for Ordinal Outcomes	151
5.1	The statistical model	152
5.1.1	A latent variable model	152
5.1.2	A nonlinear probability model	155
5.2	Estimation using ologit and oprobit	155
	Variable lists	156
	Specifying the estimation sample	156
	Weights	156
	Options	156
5.2.1	Example of attitudes toward working mothers	157
5.2.2	Predicting perfectly	160
5.3	Hypothesis testing with test and lrtest	161
5.3.1	Testing individual coefficients	161
5.3.2	Testing multiple coefficients	162
5.4	Scalar measures of fit using fitstat	163
5.5	Converting to a different parameterization*	164
5.6	The parallel regression assumption	165
5.7	Residuals and outliers using predict	168
5.8	Interpretation	169
5.8.1	Marginal change in y^*	170
5.8.2	Predicted probabilities	171
5.8.3	Predicted probabilities with predict	172
5.8.4	Individual predicted probabilities with prvalue	173
5.8.5	Tables of predicted probabilities with prtab	174
5.8.6	Graphing predicted probabilities with prgen	175
5.8.7	Changes in predicted probabilities	178
	Marginal change with prchange	178
	Marginal change with mfx compute	179
	Discrete change with prchange	179
	Computing discrete change for a 10-year increase in age	181

5.8.8	Odds ratios using listcoef	181
5.9	Less-common models for ordinal outcomes	184
5.9.1	Generalized ordered logit model	184
5.9.2	The stereotype model	185
5.9.3	The continuation ratio model	186
6	Models for Nominal Outcomes	189
6.1	The multinomial logit model	190
6.1.1	Formal statement of the model	193
6.2	Estimation using mlogit	193
	Variable lists	194
	Specifying the estimation sample	194
	Weights	194
	Options	194
6.2.1	Example of occupational attainment	195
6.2.2	Using different base categories	196
6.2.3	Predicting perfectly	198
6.3	Hypothesis testing of coefficients	199
6.3.1	mlogtest for tests of the MNLM	199
	Options	199
6.3.2	Testing the effects of the independent variables	200
	A likelihood-ratio test	200
	A Wald test	201
	Testing multiple independent variables	202
6.3.3	Tests for combining dependent categories	203
	A Wald test for combining outcomes	203
	Using test [category]*	204
	An LR test for combining outcomes	205
	Using constraint with lrtest*	205
6.4	Independence of irrelevant alternatives	207

	Hausman test of IIA	207
	Small and Hsiao test of IIA	208
	Conclusions regarding tests of IIA	210
6.5	Measures of fit	210
6.6	Interpretation	210
6.6.1	Predicted probabilities	210
6.6.2	Predicted probabilities with predict	211
	Using predict to compare mlogit and ologit	211
6.6.3	Individual predicted probabilities with prvalue	212
6.6.4	Tables of predicted probabilities with prtab	213
6.6.5	Graphing predicted probabilities with prgen	214
	Plotting probabilities for one outcome and two groups	215
	Graphing probabilities for all outcomes for one group	216
6.6.6	Changes in predicted probabilities	218
	Computing marginal and discrete change with prchange	219
	Marginal change with mfx compute	220
6.6.7	Plotting discrete changes with prchange and mlogview	221
6.6.8	Odds ratios using listcoef and mlogview	224
	Listing odds ratios with listcoef	224
	Plotting odds ratios	225
6.6.9	Using mlogplot*	230
6.6.10	Plotting estimates from matrices with mlogplot*	231
	Options for using matrices with mlogplot	232
	Global macros and matrices used by mlogplot	232
	Example	233
6.7	The conditional logit model	235
6.7.1	Data arrangement for conditional logit	236
6.7.2	Fitting the conditional logit model	237
	Options	237
	Example of the clogit model	238

6.7.3	Interpreting results from clogit	238
	Using odds ratios	238
	Using predicted probabilities	239
6.7.4	Fitting the multinomial logit model using clogit*	240
	Setting up the data	240
	Creating interactions	241
	Fitting the model	242
6.7.5	Using clogit to fit mixed models*	243
7	Models for Count Outcomes	245
7.1	The Poisson distribution	245
7.1.1	Fitting the Poisson distribution with the poisson command . . .	246
7.1.2	Computing predicted probabilities with prcounts	248
	Syntax	248
	Options	248
	Variables generated	248
7.1.3	Comparing observed and predicted counts with prcounts	249
7.2	The Poisson regression model	251
7.2.1	Fitting the PRM with poisson	252
	Variable lists	252
	Specifying the estimation sample	252
	Weights	252
	Options	253
7.2.2	Example of fitting the PRM	253
7.2.3	Interpretation using the rate μ	254
	Factor change in $E(y \mid x)$	254
	Percent change in $E(y \mid x)$	255
	Example of factor and percent change	255
	Marginal change in $E(y \mid x)$	256
	Example of marginal change using prchange	257

	Example of marginal change using mfx compute	257
	Discrete change in $E(y \mid x)$	257
	Example of discrete change using prchange	258
7.2.4	Interpretation using predicted probabilities	259
	Example of predicted probabilities using prvalue	259
	Example of predicted probabilities using prgen	261
	Example of predicted probabilities using prcounts	262
7.2.5	Exposure time*	264
7.3	The negative binomial regression model	266
7.3.1	Fitting the NBRM with nbreg	268
7.3.2	Example of fitting the NBRM	268
	Comparing the PRM and NBRM using estimates table	269
7.3.3	Testing for overdispersion	269
7.3.4	Interpretation using the rate μ	270
7.3.5	Interpretation using predicted probabilities	272
7.4	Zero-inflated count models	274
7.4.1	Fitting zero-inflated models with zinb and zip	276
	Variable lists	277
	Options	277
7.4.2	Example of fitting the ZIP and ZINB models	277
7.4.3	Interpretation of coefficients	278
7.4.4	Interpretation of predicted probabilities	280
	Predicted probabilities with prvalue	280
	Predicted probabilities with prgen	281
7.5	Comparisons among count models	283
7.5.1	Comparing mean probabilities	283
7.5.2	Tests to compare count models	284
	LR tests of α	285
	Vuong test of non-nested models	285

8	Additional Topics	287
8.1	Ordinal and nominal independent variables	287
8.1.1	Coding a categorical independent variable as a set of dummy variables	287
8.1.2	Estimation and interpretation with categorical independent variables	289
8.1.3	Tests with categorical independent variables	290
	Testing the effect of membership in one category versus the reference category	290
	Testing the effect of membership in two nonreference categories	291
	Testing that a categorical independent variable has no effect	292
	Testing whether treating an ordinal variable as interval loses information	293
8.1.4	Discrete change for categorical independent variables	294
	Computing discrete change with prchange	294
	Computing discrete change with prvalue	295
8.2	Interactions	295
8.2.1	Computing gender differences in predictions with interactions	296
8.2.2	Computing gender differences in discrete change with interactions	298
8.3	Nonlinear nonlinear models	299
8.3.1	Adding nonlinearities to linear predictors	299
8.3.2	Discrete change in nonlinear nonlinear models	300
8.4	Using praccum and forvalues to plot predictions	301
	Options	302
8.4.1	Example using age and age-squared	303
8.4.2	Using forvalues with praccum	305
8.4.3	Using praccum for graphing a transformed variable	306
8.4.4	Using praccum to graph interactions	307
8.5	Extending SPost to other estimation commands	309
8.6	Using Stata more efficiently	310
8.6.1	profile.do	310

8.6.2	Changing screen fonts and window preferences	311
8.6.3	Using ado-files for changing directories	311
8.6.4	me.hlp file	312
8.6.5	Scrolling in the Results window in Windows	312
8.7	Conclusions	313
A	Syntax for SPost Commands	315
A.1	brant	315
	Syntax	315
	Description	316
	Options	316
	Examples	316
	Saved results	317
A.2	fitstat	317
	Syntax	317
	Description	317
	Options	318
	Examples	318
	Saved results	320
A.3	listcoef	320
	Syntax	320
	Description	320
	Options	321
	Examples	321
	Saved results	323
A.4	mlogplot	323
	Syntax	323
	Description	324
	Options	324
	Examples	326

A.5 mlogtest 326

 Syntax 327

 Description 327

 Options 327

 Examples 327

 Saved results 330

 Acknowledgment 330

A.6 mlogview 330

 Syntax 330

 Description 331

 Dialog box controls 331

A.7 Overview of prchange, prgen, prtab, and prvalue 332

 Syntax 332

 Examples 333

A.8 praccum 334

 Syntax 334

 Description 334

 Options 334

 Examples 335

 New variables generated 336

A.9 prchange 336

 Syntax 337

 Description 337

 Options 337

 Examples 338

A.10 prcounts 339

 Syntax 339

 Description 339

 Options 339

 New variables generated 339

 Examples 340

A.11 prgen	341
Syntax	341
Description	341
Options	341
Examples	342
New variables generated	343
A.12 prtab	343
Syntax	343
Description	343
Options	344
Examples	344
A.13 prvalue	346
Syntax	346
Description	346
Options	346
Examples	347
Saved results	348
B Description of Datasets	349
B.1 binlfp2	349
B.2 couart2	350
B.3 gsskidvalue2	350
B.4 nomocc2	351
B.5 ordwarm2	352
B.6 science2	353
B.7 travel2	355
Author index	361
Subject index	363