

Preliminary draft of Table of Contents. Long and Freese, Regression Models for Categorical Dependent Variables Using Stata. Third Edition.

Contents

List of figures	xix
Preface	xxi
I General Information	1
1 Introduction	3
1.1 What is this book about?	3
1.2 Which models are considered?	4
1.3 Whom is this book for?	5
1.4 How is the book organized?	5
1.5 The SPost software	7
1.5.1 Updating Stata	8
1.5.2 Installing SPost13	9
Uninstalling SPost9	9
Installing SPost13 using search	10
Installing SPost13 using net install	12
1.5.3 Uninstalling SPost13	13
1.6 Sample do-files and datasets	13
1.6.1 Installing the spost13.do package	13
1.6.2 Using spex to load data and run examples	13
1.7 Getting help with SPost	14
1.7.1 What if an SPost command does not work?	14
1.7.2 Getting help from the authors	15
What we need in order to help you	16
1.8 Where can I learn more about the models?	16
2 Introduction to Stata	19

2.1	The Stata interface	19
2.2	Abbreviations	23
2.3	Getting help	23
2.3.1	Online help	23
2.3.2	PDF manuals	24
2.3.3	Error messages	24
2.3.4	Asking for help	24
2.3.5	Other resources	25
2.4	The working directory	25
2.5	Stata file types	26
2.6	Saving output to log files	26
2.7	Using and saving datasets	28
2.7.1	Data in Stata format	28
2.7.2	Data in other formats	29
2.7.3	Entering data by hand	29
2.8	Size limitations on datasets	30
2.9	Do-files	30
2.9.1	Adding comments	31
2.9.2	Long lines	32
2.9.3	Stopping a do-file while it is running	33
2.9.4	Creating do-files	33
2.9.5	Recommended structure for do-files	34
2.10	Using Stata for serious data analysis	35
2.11	Syntax of Stata commands	37
2.11.1	Commands	39
2.11.2	Variable lists	39
2.11.3	if and in qualifiers	40
2.11.4	Options	42
2.12	Managing data	42
2.12.1	Looking at your data	42

2.12.2	Getting information about variables	42
2.12.3	Missing values	46
2.12.4	Selecting observations	46
2.12.5	Selecting variables	47
2.13	Creating new variables	47
2.13.1	generate command	47
2.13.2	replace command	49
2.13.3	recode command	50
2.14	Labeling variables and values	51
2.14.1	Variable labels	51
2.14.2	Value labels	52
2.14.3	notes command	54
2.15	Global and local macros	54
2.16	Loops using foreach and forvalues	56
2.17	Graphics	58
2.18	A brief tutorial	68
2.19	A do-file template	74
2.20	Conclusion	75
3	Estimation, testing, and fit	77
3.1	Estimation	78
3.1.1	Stata's output for ML estimation	78
3.1.2	ML and sample size	79
3.1.3	Problems in obtaining ML estimates	79
3.1.4	Syntax of estimation commands	80
3.1.5	Variable lists	80
	Using factor-variable notation in the variable list	81
	Specifying interaction and polynomials	82
	More on factor-variable notation	84
3.1.6	Specifying the estimation sample	87
	Missing data	87

	Information about missing values	89
	Postestimation commands and the estimation sample	92
3.1.7	Weights and survey data	93
	Complex survey designs	94
3.1.8	Options for regression models	96
3.1.9	Robust standard errors	97
3.1.10	Reading the estimation output	99
3.1.11	Storing estimation results	101
	(Advanced) Saving estimates to a file	102
3.1.12	Reformatting output with estimates table	104
3.2	Testing	107
3.2.1	One-tailed and two-tailed tests	107
3.2.2	Wald and likelihood-ratio tests	108
3.2.3	Wald tests with test and testparm	109
3.2.4	LR tests with lrtest	111
	Avoiding invalid LR tests	112
3.3	Measures of fit	113
3.3.1	Syntax of fitstat	113
3.3.2	Methods and formulas used by fitstat	115
3.3.3	Example of fitstat	121
3.4	estat post-estimation commands	122
3.5	Conclusions	123
4	Methods of interpretation	125
4.1	Comparing linear and nonlinear models	125
4.2	Approaches to interpretation	128
4.2.1	Method of interpretation based on predictions	128
4.2.2	Method of interpretation using parameters	129
4.2.3	Stata and SPost commands for interpretation	129
4.3	Predictions for each observation	130
4.4	Predictions at specified values	131

- 4.4.1 Why use the m* commands instead of margins? 131
- 4.4.2 Using margins for predictions 132
 - Predictions using interaction and polynomial terms 137
 - Making multiple predictions 137
 - Predictions for groups defined by levels of categorical variables 141
- 4.4.3 (Advanced) Non-default predictions using margins 144
 - The predict() option 144
 - The expression() option 145
- 4.4.4 Tables of predictions using mtable 146
 - mtable with categorical and count outcomes 149
 - (Advanced) Combining and formatting tables using mtable . 150
- 4.5 Marginal effects: changes in predictions 153
 - 4.5.1 Marginal effects using margins 153
 - 4.5.2 Marginal effects using mtable 154
 - 4.5.3 Posting predictions and using mlincom 155
 - 4.5.4 Marginal effects using mchange 157
- 4.6 Plotting predictions 161
 - 4.6.1 Plotting predictions with marginsplot 161
 - 4.6.2 Plotting predictions using mgen 163
- 4.7 Interpretation of parameters 168
 - 4.7.1 The listcoef command 169
 - 4.7.2 Standardized coefficients 170
 - 4.7.3 Factor and percent change coefficients 173
- 4.8 Next steps 174

II Models for Specific Kinds of Outcomes 175

5 Models for binary outcomes: Estimation, testing, and fit 177

- 5.1 The statistical model 177
 - 5.1.1 A latent variable model 178
 - 5.1.2 A nonlinear probability model 180

5.2	Estimation using logit and probit commands	182
5.2.1	Example of logit model	183
5.2.2	Comparing logit and probit	185
5.2.3	(Advanced) Observations predicted perfectly	186
5.3	Hypothesis testing	189
5.3.1	Testing individual coefficients	189
5.3.2	Testing multiple coefficients	191
5.3.3	Comparing LR and Wald tests	194
5.4	Predicted probabilities, residuals, and influential observations	194
5.4.1	Predicted probabilities using predict	195
5.4.2	Residuals and influential observations using predict	198
5.4.3	Least likely observations	204
5.5	Measures of fit	206
5.5.1	Information criteria	206
5.5.2	Pseudo-R ² 's	208
5.5.3	(Advanced) Hosmer-Lemeshow statistic	209
5.6	Other commands for binary outcomes	211
5.7	Conclusion	212
6	Models for binary outcomes: Interpretation	213
6.1	Interpretation using regression coefficients	214
6.1.1	Interpretation using odds ratios	214
6.1.2	(Advanced) Interpretation using y^*	221
6.2	Marginal effects: changes in probabilities	224
6.2.1	Linked variables	227
6.2.2	Summary measures of change	227
	Marginal effects at the mean and marginal effects at rep- resentative values	228
	Average marginal effects	228
	Standard errors of marginal effects	229
6.2.3	Should you use the AME, the MEM, or the MER?	229

6.2.4	Examples of marginal effects	231
	Average marginal effects for continuous variables	233
	Average marginal effects for factor variables	236
	Summary table of average marginal effects	237
	Marginal effects for subgroups	239
	Marginal effects at means and marginal effects at representative values	239
	Marginal effects with powers and interactions	242
6.2.5	The distribution of marginal effects	244
6.2.6	(Advanced) Algorithm for computing the distribution of effects	248
6.3	Ideal types	252
6.3.1	Using local means with ideal types	255
6.3.2	Comparing ideal types with statistical tests	257
6.3.3	(Advanced) Using macros to test differences between ideal types	258
6.3.4	Marginal effects for ideal types	260
6.4	Tables of predicted probabilities	261
6.5	Second differences comparing marginal effects	266
6.6	Graphing predicted probabilities	267
6.6.1	Using marginsplot	268
6.6.2	Using mgen with the graph command	271
6.6.3	Graphing multiple predictions	273
6.6.4	Overlapping confidence intervals	277
6.6.5	Adding power terms and plotting predictions	280
6.6.6	(Advanced) Graphs with local means	281
6.7	Conclusion	286
7	JSL5 Models for ordinal outcomes	287
7.1	The statistical model	288
7.1.1	A latent variable model	288
7.1.2	A nonlinear probability model	291

7.2	Estimation using ologit and oprobit commands	292
7.2.1	Example of ordinal logit model	293
7.2.2	Predicting perfectly	296
7.3	Hypothesis testing	297
7.3.1	Testing individual coefficients	298
7.3.2	Testing multiple coefficients	299
7.4	Measures of fit using fitstat	301
7.5	(Advanced) Converting to a different parameterization	302
7.6	The parallel regression assumption	303
7.6.1	Testing the parallel regression assumption using oparallel	306
7.6.2	Testing the parallel regression assumption using brant	307
7.6.3	Caveat regarding the parallel regression assumption	308
7.7	Overview of interpretation	308
7.8	Interpreting transformed coefficients	309
7.8.1	Marginal change in y^*	309
7.8.2	Odds ratios	311
7.9	Interpretations based on predicted probabilities	314
7.10	Predicted probabilities with predict	315
7.11	Marginal effects	317
7.11.1	Plotting marginal effects	321
7.11.2	Marginal effects for a quick overview	326
7.12	Predicted probabilities for ideal types	327
7.12.1	(Advanced) Testing differences between ideal types	330
7.13	Tables of predicted probabilities	331
7.14	Plotting predicted probabilities	335
7.15	Probability plots and marginal effects	340
7.16	Less common models for ordinal outcomes	345
7.16.1	The stereotype logistic model	346
7.16.2	The generalized ordered logit model	346

7.16.3	(Advanced) Predictions without using factor-variable notation	348
7.16.4	The sequential logit model	352
7.17	Conclusions	356
8	Models for nominal outcomes	359
8.1	The multinomial logit model	360
8.1.1	Formal statement of the model	363
8.2	Estimation using the mlogit command	364
	Weights and complex samples	365
	Options	365
8.2.1	Example of multinomial logit model	365
8.2.2	Selecting different base outcomes	368
8.2.3	Predicting perfectly	371
8.3	Hypothesis testing	371
8.3.1	mlogtest for tests of the MNLN	372
8.3.2	Testing the effects of the independent variables	372
8.3.3	Tests for combining alternatives	376
8.4	Independence of irrelevant alternatives	379
8.4.1	Hausman test of IIA	381
8.4.2	Small–Hsiao test of IIA	382
8.5	Measures of fit	383
8.6	Overview of interpretation	384
8.7	Predicted probabilities with predict	385
8.8	Marginal effects	388
8.8.1	(Advanced) The distribution of marginal effects	392
8.9	Tables of predicted probabilities	396
8.9.1	(Advanced) Testing second differences	397
8.9.2	(Advanced) Predictions using local means and subsamples	400
8.10	Graphing predicted probabilities	403
8.11	Odds ratios	406
8.11.1	Listing odds ratios with listcoef	407

8.11.2	Plotting odds ratios	408
8.12	(Advanced) Additional models for nominal outcomes	417
8.12.1	Stereotype logistic regression	417
8.12.2	Conditional logit model	427
8.12.3	Multinomial probit model with IIA	437
8.12.4	Alternative-specific multinomial probit	441
8.12.5	Rank-ordered logit model	447
8.13	Conclusion	450
9	Models for count outcomes	453
9.1	The Poisson distribution	453
9.1.1	Fitting the Poisson distribution with the poisson command	455
9.1.2	Comparing observed and predicted counts with mgen	456
9.2	The Poisson regression model	459
9.2.1	Estimation using poisson	460
	Example of Poisson regression model	461
9.2.2	Factor and percent changes in $E(y x)$	462
	Example of factor and percent change	463
9.2.3	Marginal effects on $E(y x)$	465
	Examples of marginal effects	466
9.2.4	Interpretation using predicted probabilities	468
	Predicted probabilities using mtable and mchange	468
	Treating a count independent variable as a factor variable .	470
	Predicted probabilities using mgen	471
9.2.5	Comparing observed and predicted counts to evaluate model specification	472
9.2.6	(Advanced) Exposure time	475
9.3	The negative binomial regression model	477
9.3.1	Estimation using nbreg	479
	NB1 and NB2 variance functions	480
9.3.2	Example of negative binomial regression model	480

9.3.3	Testing for overdispersion	481
9.3.4	Comparing the PRM and NBRM using estimates table . . .	482
9.3.5	Robust standard errors	483
9.3.6	Interpretation using $E(y x)$	484
9.3.7	Interpretation using predicted probabilities	486
9.4	Models for truncated counts	489
9.4.1	Estimation using <code>tpoisson</code> and <code>tnbrms</code>	491
	Example of zero-truncated model	491
9.4.2	Interpretation using $E(y x)$	493
9.4.3	Predictions in the estimation sample	494
9.4.4	Interpretation using predicted rates and probabilities	495
9.5	(Advanced) The hurdle regression model	496
9.5.1	Fitting the hurdle model	497
9.5.2	Predictions in the sample	500
9.5.3	Predictions at user-specified values	501
9.5.4	Warning regarding sample specification	503
9.6	Zero-inflated count models	504
9.6.1	Estimation using <code>zinb</code> and <code>zip</code>	507
9.6.2	Example of zero-inflated models	508
9.6.3	Interpretation of coefficients	508
9.6.4	Interpretation of predicted probabilities	510
	Predicted probabilities with <code>mtable</code>	510
	Plotting predicted probabilities with <code>mgen</code>	512
9.7	Comparisons among count models	513
9.7.1	Comparing mean probabilities	514
9.7.2	Tests to compare count models	516
9.7.3	Using <code>countfit</code> to compare count models	520
9.8	Conclusions	525
A	Syntax for <code>mchangeplot</code>	527
A.1	Syntax for <code>mchangeplot</code>	527