

Contents

List of figures	ix
List of tables	xi
List of boxes	xiii
Preface	xv
Acknowledgments	xvii
1 What we can do with logistic regression	1
1.1 Questions we can answer using logistic regression	1
1.2 Ways to report results	2
1.3 Using Stata	4
2 Getting ready	7
2.1 Opening the dataset	7
2.2 Exploring the data	8
2.3 Labeling values for categorical variables	10
2.4 Saving the edited dataset	13
3 Conventional ordinary least-squares regression versus logistic regression	15
3.1 What OLS regression can tell us	15
3.2 What logistic regression can tell us	20
3.2.1 Robust and cluster-robust estimation	24
An imperfect model	25
Clustered sample design	26
4 Interpreting an odds ratio	29
4.1 What is an odds ratio?	29
4.2 Interpreting ORs as a percentage difference for binary predictors	31

5	What is wrong with ordinary least-squares regression for a binary outcome?	35
5.1	Hypothetical data	35
5.2	How does logistic regression fit better than ordinary least-squares linear regression?	37
6	Fitting and interpreting logistic regression models	39
6.1	Interpreting coefficients and odds ratios	39
6.2	Fitting logistic regression models with multiple predictors	41
6.3	Interpreting ORs for quantitative predictors	42
6.4	Selecting the right base level for categorical predictors	44
7	How well does the model fit the data?	47
7.1	Pseudo- R^2 measures of fit	47
7.2	Information criteria	52
7.3	Identifying cases that the model fits poorly	53
8	Sensitivity and specificity	55
8.1	Criteria for evaluating an analysis	55
8.2	Estimation of sensitivity and specificity	57
9	Receiver operating characteristic curves and cutpoints for screening tests	59
9.1	ROC curves	60
9.2	Comparing tests	66
10	Predictions using the margins command	69
10.1	What is better than reporting coefficients and odds ratios?	69
10.2	Data preparation	69
10.3	Estimating the ORs	71
10.4	The margins command	72
10.4.1	Estimating risk for categorical predictors	72
10.4.2	Estimating the risk for quantitative predictors	74
10.4.3	Estimating for a combination of categorical and quantitative variables	75

11	Graphic presentation using the marginsplot command	79
11.1	When and why	79
11.2	Graphs of categorical predictors that include three or more categories	79
11.3	Graphs with one quantitative predictor	86
11.4	Graphs with one quantitative and one categorical predictor	89
11.5	Graphs of a pair of categorical predictors	90
11.6	Graphs with a pair of quantitative predictors	91
12	Curve fitting with quadratic models	93
12.1	A hypothetical example of a quadratic model using OLS regression .	93
12.2	Estimating the curve (uncentered predictor)	97
12.3	Centering, collinearity, and nonessential collinearity	102
12.4	Estimating the curve (centered x)	103
12.5	Compare centered and uncentered models	105
12.6	Use of a quadratic with logistic regression	106
13	Interaction	115
13.1	Introduction	115
13.2	Interaction of a categorical and a quantitative variable using lo- gistic regression	117
13.3	Estimating and interpreting probabilities (uncentered)	119
13.4	Interaction of categorical variables	124
13.5	Interaction of quantitative variables	127
14	Running nestreg and postestimation commands	133
14.1	Nested logistic regression	134
14.2	Selected postestimation commands	138
15	Special topics	145
15.1	Collinearity and multicollinearity	145
15.1.1	Evaluating multicollinearity	146
15.2	Sample size	148
15.3	Small-sample bias	151
15.4	Relative risk	153

A Appendix	159
References	167
Author index	171
Subject index	173
11.1 Graphs with one quantitative predictor	167
11.2 Graphs of a pair of categorical predictors that include lines	171
11.3 Graphs with one quantitative predictor	173
11.4 Graphs with one quantitative predictor and one categorical predictor	173
11.5 Graphs of a pair of categorical predictors	173
11.6 Graphs with a pair of quantitative predictors	173
11.7 Fitting logistic models	173
11.8 Interpreting ORs for quantitative predictors	173
12.1 A hypothetical example of a multilevel model	173
12.2 Estimating the curve (understanding the bottom of the box)	173
12.3 Centering, collinearity, and nonessential collinearity	173
12.4 Estimating the curve (centered x)	173
12.5 Comparing control and noncontrol	173
12.6 Use of a quadratic with logistic regression	173
12.7 Criteria for analyzing an interaction	173
12.8 Estimation of sensitivity and specificity	173
12.9 Receiver operating characteristic (ROC) curves	173
12.10 ROC curves (uncentered)	173
12.11 Comparing tests	173
12.12 Interaction of categorical variables	173
12.13 Interaction of quantitative predictors	173
12.14 What is the best test to use?	173
12.15 Data preparation	173
12.16 Estimating the ORs	173
12.17 Selected estimation commands	173
12.18 The margin command	173
12.19 Estimating risk for categorical predictors	173
12.20 Collinearity and multicollinearity	173
12.21 Estimating multicollinearity	173
12.22 Estimating multicollinearity	173
12.23 Sample size	173
12.24 Small-sample bias	173
12.25 Relative risk	173