

Chapman & Hall/CRC
Statistics in the Social and Behavioral Sciences Series

Generalized Linear Models for Categorical and Continuous Limited Dependent Variables

Michael Smithson

The Australian National University, Canberra, Australia

Edgar C. Merkle

University of Missouri, Columbia, Missouri, USA



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Contents

Preface	xiii
List of Figures	xvii
List of Tables	xix
Notation	xxi
About the Authors	xxiii
1 Introduction and Overview	1
1.1 The Nature of Limited Dependent Variables	1
1.2 Overview of GLMs	2
1.2.1 Definition	2
1.2.2 Extensions	4
1.3 Estimation Methods and Model Evaluation	6
1.3.1 Model Evaluation and Diagnosis	6
1.3.2 Model Selection and Interpretation Issues	9
1.4 Organization of This Book	12
I Discrete Variables	15
2 Binary Variables	17
2.1 Logistic Regression	18
2.2 The Binomial GLM	20
2.2.1 Latent Variable Interpretation	22
2.2.2 Interpretation of Coefficients	22
2.2.3 Example	25
2.2.4 Extension to $n > 1$	27
2.2.5 Alternative Link Functions	30
2.3 Estimation Methods and Issues	32
2.3.1 Model Evaluation and Diagnostics	32
2.3.2 Overdispersion	38
2.3.3 Relationships to Other Models	39
2.4 Analyses in R and Stata	40

2.4.1	Analyses in R	40
2.4.2	Analyses in Stata	45
2.5	Exercises	49
3	Nominal Polytomous Variables	51
3.1	Multinomial Logit Model	51
3.2	Conditional Logit and Choice Models	58
3.3	Multinomial Processing Tree Models	61
3.4	Estimation Methods and Model Evaluation	66
3.4.1	Estimation Methods and Model Comparison	66
3.4.2	Model Evaluation and Diagnosis	67
3.5	Analyses in R and Stata	71
3.5.1	Analyses in R	71
3.5.2	Analyses in Stata	77
3.6	Exercises	80
4	Ordinal Categorical Variables	81
4.1	Modeling Ordinal Variables: Common Practice versus Best Practice	81
4.2	Ordinal Model Alternatives	82
4.2.1	The Proportional Odds Assumption	82
4.2.2	Modeling Relative Probabilities	83
4.3	Cumulative Models	84
4.3.1	The Proportional Odds Model	84
4.3.2	Example	86
4.4	Adjacent Models	89
4.4.1	The Adjacent Categories Model	89
4.4.2	Example	90
4.5	Stage Models	91
4.5.1	The Continuation Ratio Model	92
4.5.2	Example	94
4.6	Estimation Methods and Issues	97
4.6.1	Model Choice	98
4.6.2	Model Diagnostics	98
4.7	Analyses in R and Stata	102
4.7.1	Analyses in R	102
4.7.2	Analyses in Stata	107
4.8	Exercises	111
5	Count Variables	113
5.1	Distributions for Count Data	113
5.2	Poisson Regression Models	116
5.2.1	Model Definition	116
5.2.2	Example	118
5.2.3	Exposure	119
5.2.4	Overdispersion and Quasi-Poisson Models	120

5.3	Negative Binomial Models	123
5.3.1	Model Definition	123
5.3.2	Example	125
5.4	Truncated and Censored Models	125
5.5	Zero-Inflated and Hurdle Models	127
5.5.1	Hurdle Models	127
5.5.2	Zero-Inflated Models	130
5.6	Estimation Methods and Issues	133
5.6.1	Negative Binomial Model Estimation	133
5.6.2	Model Diagnostics	133
5.7	Analyses in R and Stata	137
5.7.1	Analyses in R	138
5.7.2	Analyses in Stata	144
5.8	Exercises	148
II Continuous Variables		151
6	Doubly Bounded Continuous Variables	153
6.1	Doubly Bounded versus Censored	153
6.2	The beta GLM	153
6.3	Modeling Location and Dispersion	159
6.3.1	Judged Probability of Guilt	160
6.3.2	Reading Accuracy for Dyslexic and Non-Dyslexic Readers	162
6.3.3	Model Comparison	164
6.4	Estimation Methods and Issues	166
6.4.1	Estimator Bias	168
6.4.2	Model Diagnostics	171
6.5	Zero- and One-Inflated Models	176
6.6	Finite Mixture Models	178
6.6.1	Car Dealership Example	180
6.7	Analyses in R and Stata	182
6.7.1	Analyses in R	182
6.7.2	Analyses in Stata	186
6.8	Exercises	191
7	Censoring and Truncation	193
7.1	Models for Censored and Truncated Variables	193
7.1.1	Tobit Models	197
7.2	Non-Gaussian Censored Regression	203
7.3	Estimation Methods, Model Comparison, and Diagnostics	208
7.4	Extensions of Censored Regression Models	211
7.4.1	Proportional Hazard and Proportional Odds Models	212
7.4.2	Double and Interval Censoring	214
7.4.3	Censored Quantile Regression	220

7.5	Analyses in R and Stata	222
7.5.1	Analyses in R	222
7.5.2	Analyses in Stata	228
7.6	Exercises	232
8	Extensions	235
8.1	Extensions and Generalizations	235
8.2	Multilevel Models	236
8.2.1	Multilevel Binary Logistic Regression	236
8.2.2	Multilevel Count Models	239
8.2.3	Multilevel Beta Regression	241
8.3	Bayesian Estimation	245
8.3.1	Bayesian Binomial GLM	247
8.3.2	Bayesian Beta Regression	250
8.3.3	Modeling Random Sums	253
8.4	Evaluating Relative Importance of Predictors in GLMs	256
	References	261
	Author Index	275
	Subject Index	281