
Contents

<i>Preface</i>	xiii
1 The Data Sets	1
1.1 Introduction	1
1.1.1 The School Survey on Crime and Safety	2
1.1.2 The Framingham Heart Study	2
1.1.3 Fire-Climate Interactions in the American West	2
1.1.4 English Wikipedia Clickstream Data	3
1.2 Exploratory Data Analysis	3
1.3 Gauss-Markov Assumptions	4
1.4 Data Summaries and Tables	4
1.5 Graphical Representations	4
1.5.1 Histograms	5
1.5.2 Q-Q Plots	5
1.5.3 Box-Whisker Plots	5
1.5.4 Scatter Plots	6
1.6 Pairwise Correlation	7
1.7 Machine Learning Pattern Recognition	7
1.8 Exploring the Data Sets	8
1.8.1 School Survey on Crime and Safety Data	8
1.8.2 Framingham Heart Study Data	13
1.8.3 Fire-Climate Interactions in the American West Data	17
1.8.4 English Wikipedia Clickstream Data	20
1.9 Summary	23
1.10 Further Reading	24
2 The Model-Building Process	25
2.1 Introduction	25
2.2 The Model-Building Process	26
2.2.1 Exploratory Data Analysis	26
2.2.2 Model Construction	27
2.2.3 Model Fit Diagnostics	28
2.2.4 Model Effects Analysis	28
2.2.5 Model Interpretation and Prediction	29
2.2.6 Effects and Predictive Model Differences	29
2.3 Constant Variance Response Models	30
2.4 Nonconstant Variance Response Models	31

2.5	Discrete, Categorical Response Models	32
2.6	Count Response Models	34
2.7	Time-to-Event Response Models	37
2.8	Longitudinal Response Models	39
2.9	Structural Equation Modeling	41
2.10	Effect Size	43
2.11	Model Fit Measures	43
	2.11.1 Measures of Fit	43
	2.11.2 Residual Analyses	45
2.12	Summary	48
2.13	Further Reading	49
3	Constant Variance Response Models	50
3.1	Introduction	50
3.2	School Survey on Crime and Safety	50
3.3	Framingham Heart Study	52
3.4	Fire-Climate Interactions in the American West	53
3.5	English Wikipedia Clickstream Data	55
3.6	Summary	56
3.7	Further Reading	56
4	Nonconstant Variance Response Models	57
4.1	Heterogeneity in Response Variance	57
4.2	Detecting Heteroscedasticity	58
	4.2.1 Descriptive Statistics	58
	4.2.2 Tests for Grouped Data	58
	4.2.3 Tests for Continuous Predictors	59
4.3	Variance-Stabilizing Transformations	59
	4.3.1 Selecting the Transformation	59
	4.3.2 Model Diagnostics	59
4.4	Weighted Least Squares	60
	4.4.1 WLS Estimation	60
	4.4.2 Selecting the Weights	60
4.5	SSOCS Analysis: Annual Suspensions	61
	4.5.1 Exploratory Data Analysis	61
	4.5.2 Normal Linear Model	63
	4.5.3 Outcome Transformations	63
	4.5.4 Weighted Least Squares	65
	4.5.5 Parameter Interpretations	68
	4.5.6 Model Prediction	69
4.6	Fire-Climate Analysis: Decade Averages	70
	4.6.1 Exploratory Data Analysis	70
	4.6.2 Normal Linear Model	71
	4.6.3 Weighted Least Squares	72
	4.6.4 Parameter Interpretations	74
	4.6.5 Model Prediction	74
4.7	Summary	75
4.8	Further Reading	75

Contents

ix

5	Discrete, Categorical Response Models	76
5.1	Categorical Responses	76
5.2	Binary Logistic Regression	76
	5.2.1 Descriptive Statistics for Binary Outcomes	77
	5.2.2 The Logistic Regression Model	78
	5.2.3 Interpreting Model Coefficients	78
	5.2.4 Model Fit	79
5.3	Nominal Multinomial Models	81
5.4	Ordinal Multinomial Models	82
	5.4.1 Cumulative Logit Model	83
	5.4.2 Adjacent Categories Model	83
	5.4.3 Continuation Ratio Model	84
5.5	FHS Analysis: Probability of Hypertension	85
	5.5.1 Exploratory Data Analyses	85
	5.5.2 Logistic Regression Model	86
	5.5.3 Logistic Regression Model Fit	87
	5.5.4 Model Parameter Interpretations	89
	5.5.5 Model Prediction	90
5.6	SSOCS Analysis: Probability of Bullying	93
	5.6.1 Exploratory Data Analysis	93
	5.6.2 Ordinal Multinomial Model	94
	5.6.3 Ordinal Multinomial Model Fit	96
	5.6.4 Model Parameters Interpretations	97
	5.6.5 Model Prediction	99
5.7	Clickstream Analysis: Probability of Redlink	101
	5.7.1 Exploratory Data Analysis	102
	5.7.2 Logistic Regression Model	102
	5.7.3 Logistic Regression Model Fit	103
	5.7.4 Model Parameter Interpretations	104
	5.7.5 Model Prediction	105
5.8	Summary	106
5.9	Further Reading	107
6	Count Response Models	108
6.1	Introduction	108
6.2	Modeling Count Data	109
	6.2.1 Poisson Models	109
	6.2.2 Overdispersion	110
	6.2.3 Coefficient Interpretations	111
	6.2.4 Negative Binomial Models	113
	6.2.5 Zero-Inflated Models	114
	6.2.6 Zero-Deflated Models	114
	6.2.7 Hurdle Models	115
6.3	Fire-Climate Analysis: Decade Counts	115
	6.3.1 Exploratory Data Analysis	115
	6.3.2 Poisson Model	116
	6.3.3 Negative Binomial Models	118
	6.3.4 Zero-Inflated NB Models	119

6.4	SSOCS Analysis: Annual Suspensions	123
6.4.1	Hurdle Negative Binomial Model	123
6.4.2	Model Fit	124
6.4.3	Model Interpretations	124
6.5	Clickstream Analysis: Site Pairings	126
6.5.1	Exploratory Data Analysis	126
6.5.2	Left-truncated Count Model	126
6.5.3	Count Model Fit	128
6.5.4	Coefficient Interpretations	129
6.6	Summary	130
6.7	Further Reading	131
7	Time-to-Event Response Models	132
7.1	Time-to-Event Data	132
7.2	Time-to-Event Models	133
7.3	FHS Analysis: Time to Hypertension	135
7.3.1	Life Tables	135
7.3.2	Kaplan-Meier Method	138
7.3.3	Cox Proportional Hazards Models	140
7.3.4	Time-Dependent Cox Models	145
7.4	Summary	150
7.5	Further Reading	150
8	Longitudinal Response Models	152
8.1	Longitudinal Data	152
8.2	Autocorrelation in Longitudinal Data	153
8.2.1	Descriptive Analysis	153
8.2.2	Scatter plots	153
8.2.3	Autocorrelation Plots	154
8.2.4	Variograms	155
8.2.5	Modeling Longitudinal Data	156
8.3	Marginal Models	156
8.3.1	Generalized Estimating Equations	157
8.3.2	Working Correlation Structure	157
8.3.3	Marginal Model Fit	159
8.4	Conditional Models	160
8.4.1	Random-Intercept Models	160
8.4.2	Random-Slopes Models	161
8.4.3	Conditional Model Fit	162
8.5	FHS Analysis: Probability of Hypertension	163
8.5.1	Exploratory Data Analysis	163
8.5.2	Marginal Longitudinal Model	166
8.5.3	Examining the Autocorrelation	166
8.5.4	Marginal Longitudinal Model Fit	168
8.5.5	Model Parameter Interpretations	168
8.5.6	Model Prediction	170
8.6	Fire-Climate Analysis: Decade Counts	172
8.6.1	Exploratory Data Analysis	172

Contents

xi

8.6.2	Autocorrelation in Decade Counts	175
8.6.3	Conditional Models for Decade Counts	175
8.6.4	Conditional Longitudinal Model Fit	176
8.6.5	Model Parameter Interpretations	178
8.6.6	Model Prediction	179
8.7	Summary	181
8.8	Further Reading	181
9	Structural Equation Modeling	183
9.1	Introduction	183
9.1.1	SEM Variable Categories	184
9.1.2	Model Types	185
9.1.3	SEM Paths	185
9.1.4	Confirmatory Factor Analysis	187
9.1.5	Evaluating Model Fit	188
9.2	FHS Analysis: Latent Stress	189
9.3	SSOCS Analysis: School Climate and Academic Success	194
9.4	Summary	201
9.5	Further Reading	201
10	Matching Data to Models	202
10.1	The Decision Process of Modeling	202
10.2	Results of Model Application	207
10.2.1	School Survey on Crime and Safety	207
10.2.2	Framingham Heart Study	208
10.2.3	Fire-Climate Interactions in the American West	208
10.2.4	English Wikipedia Clickstream	209
10.3	Perspectives on Modeling	209
	<i>Bibliography</i>	211
	<i>Index</i>	213