Contents

1 Introduction 1
1.1 Purpose 1
1.2 Background 4
1.3 Overview of the Adaptive Modeling Process 5
References 8

Part I Adaptive Regression Modeling

2 Adaptive Regression Modeling of Univariate Continuous Outcomes 11
2.1 Chapter Overview 11
2.2 The Death Rate Data 12
2.3 The Bivariate Regression Model and Its Parameter Estimates 12
2.4 Power Transformed Predictors 13
2.5 Cross-Validation 14
2.5.1 PRESS Formulation 15
2.5.2 PRESS Assessment of the Death Rate as a Function of the Nitric Oxide Pollution Index . . . 15
2.5.3 Formulation for Other Types of Cross-Validation . . . 15
2.6 Death Rate as a Function of the Nitric Oxide Pollution Index 17
2.7 Model Comparisons 18
2.8 Choosing the Number of Cross-Validation Folds 19
2.9 Comparison to Standard Polynomial Models 21
2.10 Penalized Likelihood Criteria for Model Selection 21
2.10.1 Formulation 22
2.10.2 Adaptive Analyses Using Penalized Likelihood Criteria 23

2.11 Monotonic Models . . . 24
2.12 Comparison to Standard Fractional Polynomial Modeling . . . 25
2.13 Log Transforms . . . 28
2.13.1 Recommended Degree 2 Fractional Polynomials . . . 29
2.13.2 Limits of Fractional Polynomials . . . 30
2.14 Impact of the Intercept . . . 30
2.15 Impact of Bounding the Nitric Oxide Pollution Index . . . 31
2.16 Death Rate as a Function of Other Predictors . . . 33
2.17 The Multiple Regression Model . . . 36
2.18 Residual Analysis . . . 37
2.19 Modeling Variances as well as Means . . . 37
2.19.1 Formulation . . . 38
2.19.2 Analysis of Death Rate Means and Variances . . . 39
2.19.3 Analysis of Means and Variances for the Simulated Data . . . 39
2.20 Overview of Analyses of Death Rates . . . 39
2.21 Overview of Analyses of the Simulated Outcome . . . 41
2.22 Chapter Summary . . . 41
References . . . 43

3 Adaptive Regression Modeling of Univariate Continuous Outcomes in SAS . . . 45
3.1 Chapter Overview . . . 45
3.2 Loading in the Death Rate Data . . . 46
3.3 Adaptive Models Based on NOindex . . . 46
3.4 Setting the Number of Cross-Validation Folds . . . 48
3.5 Standard Polynomial Models in NOindex . . . 49
3.6 Selecting Models in NOindex Using Penalized Likelihood Criteria . . . 49
3.7 Monotonic Model in NOindex . . . 50
3.8 Recommended Fractional Polynomials in NOindex . . . 50
3.9 Impact of the Log Transform of NOindex . . . 52
3.10 Zero-Intercept Models in NOindex . . . 52
3.11 Models Bounding the Impact of NOindex . . . 53
3.12 Models in Other Available Predictors . . . 54
3.13 Residual Analysis . . . 56
3.14 Modeling Variances as Well as Means . . . 59
3.15 Practice Exercises . . . 60
References . . . 62

4 Adaptive Regression Modeling of Multivariate Continuous Outcomes . . . 63
4.1 Chapter Overview . . . 63
4.2 The Dental Measurement Data . . . 64

4.3 The Marginal Multivariate Regression Model and Its Parameter Estimates . . . 64
4.3.1 Complete Data . . . 65
4.3.2 Incomplete Data . . . 65
4.3.3 Marginal Maximum Likelihood Modeling of Dependence . . . 66
4.4 LCV for Marginal Models . . . 67
4.4.1 LCV Formulation . . . 67
4.4.2 LCV Ratio Tests . . . 68
4.5 Marginal Order 1 Autoregressive Modeling of the Dental Measurement Data . . . 69
4.5.1 Order 1 Autoregressive Correlations . . . 69
4.5.2 Setting the Number of Cross-Validation Folds . . . 70
4.5.3 Moderation of the Effect of Age by Gender . . . 70
4.5.4 Geometric Combinations . . . 72
4.6 General Power Transforms . . . 73
4.6.1 Formulation . . . 74
4.6.2 The Royston and Sauerbrei Approach . . . 75
4.7 Transition Modeling of Dependence . . . 75
4.7.1 Formulation Using Averages of Prior Outcome Measurements . . . 76
4.7.2 Transition Model Induced by the Marginal AR1 Model with Constant Means . . . 78
4.7.3 Using Weighted Averages of Prior Outcome Measurements . . . 78
4.8 Transition Modeling of the Dental Measurement Data . . . 79
4.8.1 Using the Prior Dental Measurement . . . 79
4.8.2 Comparison to the Marginal Model with Exchangeable Correlations . . . 79
4.8.3 Using Multiple Prior Dental Measurements . . . 80
4.8.4 Transition Model Selection with Penalized Likelihood Criteria . . . 82
4.9 General Conditional Modeling of Dependence . . . 83
4.9.1 Formulation . . . 84
4.9.2 Conditional Models Induced by Marginal Models . . . 85
4.10 General Conditional Modeling of the Dental Measurement Data . . . 86
4.11 Adaptive GEE-Based Modeling of Multivariate Continuous Outcomes . . . 87
4.11.1 Formulation . . . 88
4.11.2 Adaptive GEE-Based Modeling of the Dental Measurement Data . . . 89
4.11.3 Assessment of the Quasi-Likelihood Information Criterion . . . 90

4.12 Analysis of the Exercise Data . . . 92
4.13 LCV with Measurement-Wise Deletion . . . 94
4.14 Revised Analysis of the Exercise Data . . . 95
4.15 Modeling Variances as Well as Means . . . 96
4.15.1 Formulation . . . 96
4.15.2 Analysis of Dental Measurement Means and Variances . . . 97
4.15.3 Transition Modeling of Strength Measurement Means with Adjusted Variances . . . 99
4.15.4 Analysis of Strength Measurement Means and Variances . . . 99
4.16 Overview of Analyses of Dental Measurements . . . 102
4.17 Overview of Analyses of Strength Measurements . . . 104
4.18 Chapter Summary . . . 105
References . . . 108

5 Adaptive Regression Modeling of Multivariate Continuous Outcomes in SAS . . . 109
5.1 Chapter Overview . . . 109
5.2 Loading the Dental Measurement Data . . . 110
5.3 Marginal Modeling of Means for the Dental Measurement Data . . . 111
5.3.1 Marginal Models of Mean Dental Measurement in Age of the Child . . . 111
5.3.2 Marginal Moderation Models of Mean Dental Measurement in Age and Gender of the Child . . . 113
5.3.3 Residual Analysis of the Adaptive Marginal Moderation Model . . . 117
5.4 Conditional Modeling of Means for the Dental Measurement Data . . . 121
5.4.1 Transition Models for Mean Dental Measurement . . . 121
5.4.2 Residual Analysis of the Adaptive Transition Model . . . 122
5.4.3 General Conditional Models for Mean Dental Measurement . . . 124
5.5 Analyzing the Exercise Data . . . 125
5.6 Modeling Variances as Well as Means . . . 125
5.6.1 Marginal Models for Dental Measurements . . . 125
5.6.2 Transition Models for Dental Measurements . . . 127
5.6.3 Clock Time Assessments . . . 128
5.7 Practice Exercises . . . 129
References . . . 132

6 Adaptive Transformation of Positive Valued Continuous Outcomes . . . 133
6.1 Chapter Overview . . . 133
6.2 Transformation of the Outcome Variable . . . 133
6.3 Formulation for Power-Adjusted Likelihoods and LCV Scores . . . 134
6.3.1 Univariate Outcomes . . . 134
6.3.2 Multivariate Outcomes . . . 135
6.4 Analyses of Transformed Death Rates . . . 137
6.5 Analyses of the Transformed Simulated Outcome . . . 137
6.6 Analyses of Transformed Dental Measurements . . . 138
6.7 Analyses of Transformed Strength Measurements . . . 139
6.8 The Plasma Beta-Carotene Data . . . 140
6.9 Analyses of Untransformed Plasma Beta-Carotene Levels . . . 141
6.10 Analyses of Transformed Plasma Beta-Carotene Levels . . . 142
6.11 Overview of Analyses of Death Rates . . . 143
6.12 Overview of Analyses of the Simulated Outcome . . . 144
6.13 Overview of Analyses of Dental Measurements . . . 144
6.14 Overview of Analyses of Strength Measurements . . . 144
6.15 Overview of Analyses of Plasma Beta-Carotene Levels . . . 144
6.16 Chapter Summary . . . 145
References . . . 146

7 Adaptive Transformation of Positive Valued Continuous Outcomes in SAS . . . 149
7.1 Chapter Overview . . . 149
7.2 Loading in the Plasma Beta-Carotene Data . . . 149
7.3 Adaptive Transformation of Plasma Beta-Carotene Levels . . . 150
7.4 Adaptive Transformation of Dental Measurements . . . 153
7.4.1 Using Transition Models . . . 153
7.4.2 Using Marginal Models . . . 154
7.5 Practice Exercises . . . 155
References . . . 157

Part II Adaptive Logistic Regression Modeling

8 Adaptive Logistic Regression Modeling of Univariate Dichotomous and Polytomous Outcomes . . . 161
8.1 Chapter Overview . . . 161
8.2 The Mercury Level Data . . . 162
8.3 Multiple Logistic Regression Modeling of Dichotomous Outcomes . . . 162
8.3.1 Multiple Logistic Regression Model Formulation . . . 162
8.3.2 Odds Ratio Function Formulation . . . 163

8.4 Dichotomous Mercury Level as a Function of Weight 164
8.5 Dichotomous Mercury Level as a Function of Length 165
8.6 Dichotomous Mercury Level as a Function of Weight and Length 167
8.7 Multiple Logistic Regression Modeling of Polytomous Outcomes 169
8.7.1 Multinomial Regression 169
8.7.2 Ordinal Regression 171
8.8 Mercury Level Categorized into Three Ordinal Levels 172
8.9 Polytomous Mercury Level as a Function of Weight 172
8.10 Polytomous Mercury Level as a Function of Length 173
8.11 Polytomous Mercury Level as a Function of Weight and Length 175
8.12 Proportion of Correct Deleted Predictions 176
8.12.1 Formulation 177
8.12.2 Example Analyses of Dichotomous Mercury Level 177
8.12.3 Example Analyses of Polytomous Mercury Level 178
8.13 Modeling Dispersions as Well as Means 178
8.13.1 Formulation for Dichotomous Outcomes 179
8.13.2 Formulation for Polytomous Outcomes 180
8.13.3 Analysis of Dichotomous Mercury Level Means and Dispersions 182
8.14 Overview of Analyses of Dichotomous Mercury Levels 183
8.15 Overview of Analyses of Polytomous Mercury Levels 184
8.16 Chapter Summary 185
References 186

9 Adaptive Logistic Regression Modeling of Univariate Dichotomous and Polytomous Outcomes in SAS 187
9.1 Chapter Overview 187
9.2 Loading in the Mercury Level Data 188
9.3 Modeling Means for Merchigh Based on Weight 189
9.4 Modeling Means for Merchigh Based on Length 193
9.5 Grouped Residuals for Univariate Dichotomous Outcomes 194
9.6 Grouped Residual Analysis of Merchigh as a Function of Length 196
9.7 Modeling Means for Merchigh Based on Weight and Length 198
9.8 Modeling Means for Merclevel Based on Weight and Length 198
9.9 Grouped Residuals for Univariate Polytomous Outcomes . . . 201
9.9.1 Multinomial Regression 201
9.9.2 Ordinal Regression 202

9.10 Grouped Residual Analysis of Merclevel as a Function of Length 203
9.11 Modeling Dispersions as Well as Means for the Dichotomous Outcome Merchigh 204
9.12 Modeling Dispersions as Well as Means for the Polytomous Outcome Merclevel 206
9.13 Practice Exercises 211
References 212

10 Adaptive Logistic Regression Modeling of Multivariate Dichotomous and Polytomous Outcomes 213
10.1 Chapter Overview 213
10.2 The Respiratory Status Data 214
10.3 Conditional Modeling of Multivariate Dichotomous Outcomes 215
10.3.1 Conditional Modeling of Means Assuming Unit Dispersions 215
10.3.2 Conditional Modeling of Dispersions as Well as Means 217
10.4 Transition Modeling of Post-Baseline Dichotomous Respiratory Status 218
10.4.1 Unit Dispersions Models 218
10.4.2 Non-Unit Dispersions Models 221
10.5 Conditional Modeling of Multivariate Polytomous Outcomes 223
10.6 Transition Modeling of Post-Baseline Polytomous Respiratory Status 224
10.6.1 Unit Dispersions Models 224
10.6.2 Non-Unit Dispersions Models 226
10.7 Adaptive GEE-Based Modeling of Multivariate Dichotomous and Polytomous Outcomes 227
10.7.1 Dichotomous Outcomes 227
10.7.2 Polytomous Outcomes 231
10.7.3 Comparing Transition Models to Marginal GEE-Based Models 233
10.8 Adaptive GEE-Based Modeling of Post-Baseline Respiratory Status 234
10.9 Overview of Analyses of Post-Baseline Dichotomous Respiratory Status 235
10.10 Overview of Analyses of Post-Baseline Polytomous Respiratory Status 236
10.11 Chapter Summary 237
References 237

11 Adaptive Logistic Regression Modeling of Multivariate Dichotomous and Polytomous Outcomes in SAS 239
11.1 Chapter Overview 239
11.2 Loading in the Respiratory Status Data 239
11.3 Transition Modeling of Dichotomous Respiratory Status 242
11.4 Marginal GEE-Based Modeling of Dichotomous Respiratory Status 246
11.5 Modeling of Polytomous Outcomes 248
11.6 Practice Exercises 249
References 251

Part III Adaptive Poisson Regression Modeling

12 Adaptive Poisson Regression Modeling of Univariate Count Outcomes 255
12.1 Chapter Overview 255
12.2 The Skin Cancer Data 255
12.3 Multiple Poisson Regression Modeling of Count Outcomes 256
12.3.1 Unit Dispersions Formulation 256
12.3.2 Non-Unit Dispersions Formulation 257
12.4 Skin Cancer Rates as a Function of the Minimum Age and City of Residence 258
12.4.1 Modeling Means for Skin Cancer Rates with Constant Dispersions Models 258
12.4.2 Modeling Dispersions as Well as Means for Skin Cancer Rates 260
12.5 Overview of Analyses of Skin Cancer Rates 261
12.6 Chapter Summary 262
References 263

13 Adaptive Poisson Regression Modeling of Univariate Count Outcomes in SAS 265
13.1 Chapter Overview 265
13.2 Loading in the Skin Cancer Data 265
13.3 Modeling Means for Skin Cancer Rates 266
13.4 Modeling Dispersions as Well as Means for Skin Cancer Rates 269
13.5 Practice Exercises 273
References 274

14 Adaptive Poisson Regression Modeling of Multivariate Count Outcomes 275
14.1 Chapter Overview 275
14.2 The Epileptic Seizures Data 276

14.3 Conditional Modeling of Multivariate Count Outcomes 276
14.3.1 Conditional Modeling of Means Assuming Unit Dispersions 277
14.3.2 Conditional Modeling of Dispersions as Well as Means 278
14.4 Transition Modeling of Post-Baseline Seizure Rates 280
14.4.1 Constant Dispersions Models 280
14.4.2 Non-Constant Dispersions Models 282
14.5 Adaptive GEE-Based Modeling of Multivariate Count Outcomes 282
14.6 Adaptive GEE-Based Modeling of Post-Baseline Seizure Rates 283
14.7 Overview of Analyses of Post-Baseline Seizure Rates 284
14.8 Chapter Summary 284
References 285

15 Adaptive Poisson Regression Modeling of Multivariate Count Outcomes in SAS 287
15.1 Chapter Overview 287
15.2 Loading in the Epileptic Seizures Data 287
15.3 Transition Modeling of Post-Baseline Seizure Rates 289
15.4 Marginal GEE-Based Modeling of Post-Baseline Seizure Rates 293
15.5 Practice Exercises 294
Reference 295

Part IV Alternative Nonparametric Regression Modeling

16 Generalized Additive Modeling 299
16.1 Chapter Overview 299
16.2 Formulation of GAMs for Univariate Continuous Outcomes 300
16.3 Formulation of Likelihood Cross-Validation for GAMs 301
16.4 Other Forms of Cross-Validation 302
16.5 GAM Analyses of Deathrate as a Function of the Nitric Oxide Pollution Index 302
16.6 GAM Analyses of Deathrate as a Function of Other Singleton Predictors 305
16.7 GAM Analyses of Deathrate as a Function of Two Predictors 306
16.8 GAM Analyses of the Full Deathrate Data 308
16.9 Formulation of GAMs for Dichotomous Outcomes 310
16.10 GAM Analyses of the Mercury Level Data 311
16.11 Overview of Analyses of Death Rates 312
16.12 Overview of Analyses of Dichotomous Mercury Levels 313
16.13 Chapter Summary 313
References 314

17 Generalized Additive Modeling in SAS 315
17.1 Chapter Overview 315
17.2 Invoking PROC GAM 315
17.3 Generating LCV Scores for GAMs 317
17.4 Multiple Predictor GAMs 319
17.5 GAMs for Dichotomous Outcomes 323
17.6 Practice Exercises 326
References 327

18 Multivariate Adaptive Regression Spline Modeling 329
18.1 Chapter Overview 329
18.2 Description of MARS Modeling 330
18.3 MARS Analyses of Death Rates 330
18.3.1 MARS Analyses Based on NObnded 330
18.3.2 MARS Analyses Based on Rain 331
18.3.3 MARS Analyses Based on NObnded and Rain 331
18.3.4 MARS Analyses Based on the Full Set of Available Predictors 332
18.4 MARS Analyses of the Mercury Level Data 332
18.4.1 MARS Analyses Based on Weight of Fish 333
18.4.2 MARS Analyses Based on Length of Fish 333
18.4.3 MARS Analyses Based on Weight and Length of Fish 334
18.5 Overview of MARS Analyses of Death Rates 335
18.6 Overview of MARS Analyses of Dichotomous Mercury Levels 336
18.7 Chapter Summary 337
References 338

19 Multivariate Adaptive Regression Spline Modeling in SAS 339
19.1 Chapter Overview 339
19.2 Invoking PROC ADAPTIVEREG 339
19.3 Generating LCV Scores for MARS Models 341
19.4 Multiple Predictor MARS Models 342
19.5 MARS Models for Dichotomous Outcomes 343
19.6 Practice Exercises 348
Reference 349

Part V The Adaptive Regression Modeling Process

20 Adaptive Regression Modeling Formulation 353
20.1 Chapter Overview 353
20.2 Overview of General Regression Modeling Formulation 353
20.3 Overview of Model Selection Approaches 355
20.3.1 Using Cross-Validation Based on Likelihood or Likelihood-Like Functions 355
20.3.2 Alternate Model Selection Approaches 357

20.4 The Adaptive Modeling Process 358
20.4.1 Conditional Predictors 358
20.4.2 Power Transforms 359
20.4.3 Selecting a Power for a Primary Predictor 359
20.4.4 Adjusting the Transforms of a Base Model 361
20.4.5 Expanding a Base Model 362
20.4.6 Considering Geometric Combinations 363
20.4.7 Contracting a Base Model 365
20.4.8 Tolerance Parameter Settings 366
20.4.9 The Complete Adaptive Model Selection Process . . 367
20.4.10 Computing Transforms 368
20.4.11 Avoiding Redundant Transforms 369
References 370

Index 371