Contents

Preface xi

Chapter 1
Varieties of Count Data 1

Some Points of Discussion 1
1.1 What Are Counts? 1
1.2 Understanding a Statistical Count Model 3
1.2.1 Basic Structure of a Linear Statistical Model 3
1.2.2 Models and Probability 7
1.2.3 Count Models 9
1.2.4 Structure of a Count Model 16
1.3 Varieties of Count Models 18
1.4 Estimation – the Modeling Process 22
1.4.1 Software for Modeling 22
1.4.2 Maximum Likelihood Estimation 23
1.4.3 Generalized Linear Models and IRLS Estimation 31
1.5 Summary 33

Chapter 2
Poisson Regression 35

Some Points of Discussion 35
2.1 Poisson Model Assumptions 36
2.2 Apparent Overdispersion 39

2.3 Constructing a "True" Poisson Model 41
2.4 Poisson Regression: Modeling Real Data 48
2.5 Interpreting Coefficients and Rate Ratios 55
2.5.1 How to Interpret a Poisson Coefficient and Associated Statistics 55
2.5.2 Rate Ratios and Probability 59
2.6 Exposure: Modeling over Time, Area, and Space 62
2.7 Prediction 66
2.8 Poisson Marginal Effects 68
2.8.1 Marginal Effect at the Mean 69
2.8.2 Average Marginal Effects 70
2.8.3 Discrete Change or Partial Effects 71
2.9 Summary 73

Chapter 3
Testing Overdispersion **74**

Some Points of Discussion 74
3.1 Basics of Count Model Fit Statistics 74
3.2 Overdispersion: What, Why, and How 81
3.3 Testing Overdispersion 81
3.3.1 Score Test 84
3.3.2 Lagrange Multiplier Test 87
3.3.3 Chi2 Test: Predicted versus Observed Counts 88
3.4 Methods of Handling Overdispersion 92
3.4.1 Scaling Standard Errors: Quasi-count Models 92
3.4.2 Quasi-likelihood Models 96
3.4.3 Sandwich or Robust Variance Estimators 99
3.4.4 Bootstrapped Standard Errors 105
3.5 Summary 106

Chapter 4
Assessment of Fit **108**

Some Points of Discussion 108
4.1 Analysis of Residual Statistics 108
4.2 Likelihood Ratio Test 112

4.2.1 Standard Likelihood Ratio Test 112
4.2.2 Boundary Likelihood Ratio Test 114
4.3 Model Selection Criteria 116
4.3.1 Akaike Information Criterion 116
4.3.2 Bayesian Information Criterion 119
4.4 Setting up and Using a Validation Sample 122
4.5 Summary and an Overview of the Modeling Process 123
4.5.1 Summary of What We Have Thus Far Discussed 124

Chapter 5
Negative Binomial Regression 126

Some Points of Discussion 126
5.1 Varieties of Negative Binomial Models 126
5.2 Negative Binomial Model Assumptions 128
5.2.1 A Word Regarding Parameterization of the Negative Binomial 133
5.3 Two Modeling Examples 136
5.3.1 Example: **rwm1984** 136
5.3.2 Example: **medpar** 148
5.4 Additional Tests 152
5.4.1 General Negative Binomial Fit Tests 152
5.4.2 Adding a Parameter – NB-P Negative Binomial 153
5.4.3 Modeling the Dispersion – Heterogeneous Negative Binomial 156
5.5 Summary 160

Chapter 6
Poisson Inverse Gaussian Regression 162

Some Points of Discussion 162
6.1 Poisson Inverse Gaussian Model Assumptions 162
6.2 Constructing and Interpreting the PIG Model 165
6.2.1 Software Considerations 165
6.2.2 Examples 165
6.3 Summary – Comparing Poisson, NB, and PIG Models 170

Chapter 7
Problems with Zeros **172**

Some Points of Discussion 172
7.1 Counts without Zeros – Zero-Truncated Models 173
7.1.1 Zero-Truncated Poisson (ZTP) 174
7.1.2 Zero-Truncated Negative Binomial (ZTNB) 177
7.1.3 Zero-Truncated Poisson Inverse Gaussian (ZTPIG) 180
7.1.4 Zero-Truncated NB-P (ZTNBP) 182
7.1.5 Zero-Truncated Poisson Log-Normal (ZTPLN) 183
7.1.6 Zero-Truncated Model Summary 184
7.2 Two-Part Hurdle Models 184
7.2.1 Poisson and Negative Binomial Logit Hurdle Models 185
7.2.2 PIG-Logit and Poisson Log-Normal Hurdle Models 192
7.2.3 PIG-Poisson Hurdle Model 194
7.3 Zero-Inflated Mixture Models 196
7.3.1 Overview and Guidelines 196
7.3.2 Fit Tests for Zero-Inflated Models 197
7.3.3 Fitting Zero-Inflated Models 197
7.3.4 Good and Bad Zeros 198
7.3.5 Zero-Inflated Poisson (ZIP) 199
7.3.6 Zero-Inflated Negative Binomial (ZINB) 202
7.3.7 Zero-Inflated Poisson Inverse Gaussian (ZIPIG) 206
7.4 Summary – Finding the Optimal Model 207

Chapter 8
Modeling Underdispersed Count Data – Generalized Poisson **210**

Some Points of Discussion 210

Chapter 9
Complex Data: More Advanced Models **217**

Types of Data and Problems Dealt with in This Chapter 217
9.1 Small and Unbalanced Data – Exact Poisson Regression 218
9.2 Modeling Truncated and Censored Counts 224
9.2.1 Truncated Count Models 225
9.2.2 Censored Count Models 229
9.2.3 Poisson-Logit Hurdle at 3 Model 231

9.3 Counts with Multiple Components – Finite Mixture Models 232
9.4 Adding Smoothing Terms to a Model – GAM 235
9.5 When All Else Fails: Quantile Count Models 238
9.6 A Word about Longitudinal and Clustered Count Models 239
9.6.1 Generalized Estimating Equations (GEEs) 239
9.6.2 Mixed-Effects and Multilevel Models 241
9.7 Three-Parameter Count Models 245
9.8 Bayesian Count Models – Future Directions of Modeling? 248
9.9 Summary 252

Appendix: SAS Code **255**

Bibliography **269**

Index **277**