

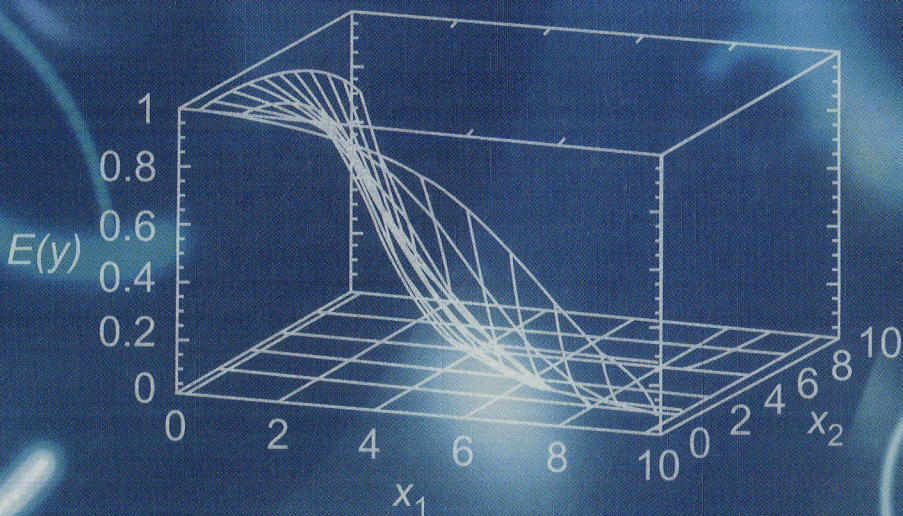
Wiley Series in Probability and Statistics

# Generalized Linear Models

*with* Applications in  
Engineering and the Sciences

*Second Edition*

Raymond H. Myers, Douglas C. Montgomery,  
G. Geoffrey Vining, and Timothy J. Robinson



# Contents

<b>Preface</b>	<b>xi</b>
<b>1. Introduction to Generalized Linear Models</b>	<b>1</b>
1.1 Linear Models, 1	
1.2 Nonlinear Models, 3	
1.3 The Generalized Linear Model, 4	
<b>2. Linear Regression Models</b>	<b>9</b>
2.1 The Linear Regression Model and Its Application, 9	
2.2 Multiple Regression Models, 10	
2.2.1 Parameter Estimation with Ordinary Least Squares, 10	
2.2.2 Properties of the Least Squares Estimator and Estimation of $\sigma^2$ , 15	
2.2.3 Hypothesis Testing in Multiple Regression, 19	
2.2.4 Confidence Intervals in Multiple Regression, 29	
2.2.5 Prediction of New Response Observations, 32	
2.2.6 Linear Regression Computer Output, 34	
2.3 Parameter Estimation Using Maximum Likelihood, 34	
2.3.1 Parameter Estimation Under the Normal-Theory Assumptions, 34	
2.3.2 Properties of the Maximum Likelihood Estimators, 38	
2.4 Model Adequacy Checking, 39	
2.4.1 Residual Analysis, 39	
2.4.2 Transformation of the Response Variable Using the Box–Cox Method, 43	

- 2.4.3 Scaling Residuals, 45
- 2.4.4 Influence Diagnostics, 50
- 2.5 Using R to Perform Linear Regression Analysis, 52
- 2.6 Parameter Estimation by Weighted Least Squares, 54
  - 2.6.1 The Constant Variance Assumption, 54
  - 2.6.2 Generalized and Weighted Least Squares, 55
  - 2.6.3 Generalized Least Squares and Maximum Likelihood, 58
- 2.7 Designs for Regression Models, 58

Exercises, 65

### 3. Nonlinear Regression Models

77

- 3.1 Linear and Nonlinear Regression Models, 77
  - 3.1.1 Linear Regression Models, 77
  - 3.1.2 Nonlinear Regression Models, 78
  - 3.1.3 Origins of Nonlinear Models, 79
- 3.2 Transforming to a Linear Model, 81
- 3.3 Parameter Estimation in a Nonlinear System, 84
  - 3.3.1 Nonlinear Least Squares, 84
  - 3.3.2 The Geometry of Linear and Nonlinear Least Squares, 86
  - 3.3.3 Maximum Likelihood Estimation, 86
  - 3.3.4 Linearization and the Gauss–Newton Method, 89
  - 3.3.5 Using R to Perform Nonlinear Regression Analysis, 99
  - 3.3.6 Other Parameter Estimation Methods, 100
  - 3.3.7 Starting Values, 101
- 3.4 Statistical Inference in Nonlinear Regression, 102
- 3.5 Weighted Nonlinear Regression, 106
- 3.6 Examples of Nonlinear Regression Models, 107
- 3.7 Designs for Nonlinear Regression Models, 108

Exercises, 111

### 4. Logistic and Poisson Regression Models

119

- 4.1 Regression Models Where the Variance Is a Function of the Mean, 119
- 4.2 Logistic Regression Models, 120
  - 4.2.1 Models with a Binary Response Variable, 120
  - 4.2.2 Estimating the Parameters in a Logistic Regression Model, 123

- 4.2.3 Interpretation of the Parameters in a Logistic Regression Model, 128
- 4.2.4 Statistical Inference on Model Parameters, 132
- 4.2.5 Lack-of-Fit Tests in Logistic Regression, 143
- 4.2.6 Diagnostic Checking in Logistic Regression, 155
- 4.2.7 Classification and the Receiver Operating Characteristic Curve, 162
- 4.2.8 A Biological Example of Logistic Regression, 164
- 4.2.9 Other Models for Binary Response Data, 168
- 4.2.10 More than Two Categorical Outcomes, 169
- 4.3 Poisson Regression, 176
- 4.4 Overdispersion in Logistic and Poisson Regression, 184
- Exercises, 189

## 5. The Generalized Linear Model

202

- 5.1 The Exponential Family of Distributions, 202
- 5.2 Formal Structure for the Class of Generalized Linear Models, 205
- 5.3 Likelihood Equations for Generalized Linear models, 207
- 5.4 Quasi-Likelihood, 211
- 5.5 Other Important Distributions for Generalized Linear Models, 213
  - 5.5.1 The Gamma Family, 214
  - 5.5.2 Canonical Link Function for the Gamma Distribution, 215
  - 5.5.3 Log Link for the Gamma Distribution, 215
- 5.6 A Class of Link Functions—The Power Function, 216
- 5.7 Inference and Residual Analysis for Generalized Linear Models, 217
- 5.8 Examples with the Gamma Distribution, 220
- 5.9 Using R to Perform GLM Analysis, 229
  - 5.9.1 Logistic Regression, Each Response is a Success or Failure, 231
  - 5.9.2 Logistic Regression, Response is the Number of Successes Out of  $n$  Trials, 232
  - 5.9.3 Poisson Regression, 232
  - 5.9.4 Using the Gamma Distribution with a Log Link, 233
- 5.10 GLM and Data Transformation, 233

- 5.11 Modeling Both a Process Mean and Process Variance Using GLM, 240
  - 5.11.1 The Replicated Case, 240
  - 5.11.2 The Unreplicated Case, 244
- 5.12 Quality of Asymptotic Results and Related Issues, 250
  - 5.12.1 Development of an Alternative Wald Confidence Interval, 250
  - 5.12.2 Estimation of Exponential Family Scale Parameter, 259
  - 5.12.3 Impact of Link Misspecification on Confidence Interval Coverage and Precision, 260
  - 5.12.4 Illustration of Binomial Distribution with a True Identity Link but with Logit Link Assumed, 260
  - 5.12.5 Poisson Distribution with a True Identity Link but with Log Link Assumed, 262
  - 5.12.6 Gamma Distribution with a True Inverse Link but with Log Link Assumed, 263
  - 5.12.7 Summary of Link Misspecification on Confidence Interval Coverage and Precision, 264
  - 5.12.8 Impact of Model Misspecification on Confidence Interval Coverage and Precision, 264

Exercises, 267

## 6. Generalized Estimating Equations

272

- 6.1 Data Layout for Longitudinal Studies, 272
- 6.2 Impact of the Correlation Matrix  $\mathbf{R}$ , 274
- 6.3 Iterative Procedure in the Normal Case, Identity Link, 275
- 6.4 Generalized Estimating Equations for More Generalized Linear Models, 277
  - 6.4.1 Structure of  $\mathbf{V}_j$ , 278
  - 6.4.2 Iterative Computation of Elements in  $\mathbf{R}$ , 283
- 6.5 Examples, 283
- 6.6 Summary, 308

Exercises, 311

## 7. Random Effects in Generalized Linear Models

319

- 7.1 Linear Mixed Effects Models, 320
  - 7.1.1 Linear Regression Models, 320
  - 7.1.2 General Linear Mixed Effects Models, 322
  - 7.1.3 Covariance Matrix,  $\mathbf{V}$ , 326

- 7.1.4 Parameter Estimation in the General Linear Mixed Model, 332
- 7.1.5 Statistical Inference on Regression Coefficients and Variance Components, 334
- 7.1.6 Conditional and Marginal Means, 337
- 7.1.7 Estimation of the Random Coefficients, 338
- 7.1.8 Examples Revisited, 340
- 7.1.9 Diagnostics, 346
- 7.2 Generalized Linear Mixed Models, 354
  - 7.2.1 The Generalized Linear Mixed Model, 357
  - 7.2.2 Parameter Estimation in the GLMM, 360
  - 7.2.3 Statistical Inference on Regression Coefficients and Variance Components, 366
  - 7.2.4 Subject-Specific Versus Population-Averaged Prediction Models, 368
  - 7.2.5 Examples Revisited, 369
  - 7.2.6 Diagnostics, 382
- 7.3 Generalized Linear Mixed Models Using Bayesian Methods, 388
  - 7.3.1 Model Formulation, 388
  - 7.3.2 Bayesian Inference, 390
  - 7.3.3 Inference on Response Distribution Characteristics, 395

Exercises, 397

## **8. Designed Experiments and the Generalized Linear Model**

**408**

- 8.1 Introduction, 408
- 8.2 Experimental Designs for Generalized Linear Models, 409
  - 8.2.1 Review of Two-Level Factorial and Fractional Factorial Designs, 409
  - 8.2.2 Finding Optimal Designs in GLMs, 411
  - 8.2.3 The Use of Standard Designs in Generalized Linear Models, 421
  - 8.2.4 Orthogonal Designs in GLM: The Variance-Stabilizing Link, 424
  - 8.2.5 Use of Other Links, 427
  - 8.2.6 Further Comments Concerning the Nature of the Design, 436
- 8.3 GLM Analysis of Screening Experiments, 437

Exercises, 458

<b>Appendix A.1.</b>	<b>Background on Basic Test Statistics</b>	<b>464</b>
<b>Appendix A.2.</b>	<b>Background from the Theory of Linear Models</b>	<b>467</b>
<b>Appendix A.3.</b>	<b>The Gauss–Markov Theorem, <math>\text{Var}(\varepsilon) = \sigma^2\mathbf{I}</math></b>	<b>472</b>
<b>Appendix A.4.</b>	<b>The Relationship Between Maximum Likelihood Estimation of the Logistic Regression Model and Weighted Least Squares</b>	<b>474</b>
<b>Appendix A.5.</b>	<b>Computational Details for GLMs for a Canonical Link</b>	<b>478</b>
<b>Appendix A.6.</b>	<b>Computational Details for GLMs for a Noncanonical Link</b>	<b>481</b>
	<b>References</b>	<b>484</b>
	<b>Index</b>	<b>493</b>