Contents

Preface **ix**

1 Examples and Limits of the GLM **1**
1.1 Motivation 1
1.2 A Review of Basic Statistical Ideas 2
1.3 GLM Definition 4
1.4 GLM Examples 4
1.5 Student Goals 5
1.6 Homework Exercises 5

2 Statement of the Model, Estimation, and Testing **7**
2.1 Motivation 7
2.2 Statement of the Model 8
2.3 Least Squares Assumptions 10
2.4 Discussion of Homogeneity 11
2.5 Gaussian Errors Assumption 13
2.6 Estimation of the GLM 14
2.7 Hypothesis Testing for the GLM 20
2.8 Homework Exercises 27

3 Some Distributions for the GLM **29**
3.1 Motivation 29
3.2 A Full-Rank Basis for Less-than-Full-Rank Models 30
3.3 Definitions and Theorems 31
3.4 GLM Distributions 32
3.5 Definitions and Properties of Residuals 36
3.6 Homework Exercises 41

4 Multiple Regression: General Considerations **43**
4.1 Motivation 43
4.2 Definitions of Basic Sums of Squares 44
4.3 The Nature of the Intercept 47
4.4 Models that Span but May Not Include an Intercept 48
4.5 Corrected *SS* 49

4.6 Intercept-Only Model 53
4.7 Null Model 55
4.8 Overall ANOVA Table for Multiple Regression 56
4.9 Usual ("Corrected") Overall Test for Regression 59
4.10 "Uncorrected" Overall Test for Regression 61
4.11 Strength of Association 62
4.12 Homework Exercises 64

5 Testing Hypotheses in Multiple Regression 67
5.1 Motivation 67
5.2 Choosing an Error Term 68
5.3 Review of GLH Concepts 69
5.4 Model Pools 69
5.5 Test Class 1: Overall 73
5.6 Test Class 2: Addition of One Variable 78
5.7 Test Class 3: Tests of the Intercept 82
5.8 Test Class 4: Addition of a Group of Variables 86
5.9 Test Class 5: GLH Tests 91
5.10 The Multiple Testing Issue 93
5.11 Interaction 94
5.12 Homework Exercises 95

6 Correlations 97
6.1 Motivation 97
6.2 Interpreting ρ^2 97
6.3 Correlation Formulas 102
6.4 Partial Correlation 103
6.5 Semipartial Correlation 106
6.6 Relating Semipartial Correlations to Standardized Regression Coefficients 108
6.7 Multiple Partial Correlation 109
6.8 Multiple Semipartial Correlation 110
6.9 Hypothesis Tests for Correlations 111
6.10 Relating Multiple Partials and Semipartials to Regression Coefficient Tests 114
6.11 Using Correlations to Interpret Added-in-Order Tests 115
6.12 Computing Partial Correlations 116
6.13 Some Useful Properties of Correlations 118
6.14 The Importance and Utility of Correlations 119
6.15 Homework Exercises 120

7 GLM Assumption Diagnostics 121
7.1 Motivation 121
7.2 The First Step: Get to Know Your Data 122
7.3 Residual Analysis 133

7.4 Outliers 146
7.5 Homework Exercises 160

8 GLM Computation Diagnostics 163
8.1 Motivation 163
8.2 Single Variable Problems and Solutions 163
8.3 Collinearity Definitions and Concepts 164
8.4 Matrices of Interest 165
8.5 Models Corresponding to Cross-Products Matrices 168
8.6 Eigenanalysis 172
8.7 R_j^2, Tolerance and VIF 177
8.8 Detecting Numerical Inaccuracy 180
8.9 Treating Regression Problems 182
8.10 Homework Exercises 183

9 Polynomial Regression 185
9.1 Motivation 185
9.2 Polynomial Models of Most Interest 186
9.3 Examples 186
9.4 Lack-of-Fit Tests 188
9.5 Limitations of Natural Polynomials 189
9.6 Orthogonal Polynomials 191
9.7 Strategies for Accurate Computations with Polynomials 195
9.8 Strategy for Choosing a Model 200
9.9 Polynomial Interactions and Response Surfaces 201
9.10 Homework Exercises 202

10 Transformations 205
10.1 Motivation 205
10.2 General Principles 206
10.3 Power Transformation of the Response 206
10.4 Other Comments 218
10.5 Transformations of Predictors 219
10.6 Pitfalls 219
10.7 Homework Exercises 220

11 Selecting the Best Model 223
11.1 Motivation 223
11.2 Overview of Solution Strategies 224
11.3 Step 1: Specify the Maximum Model 224
11.4 Step 2: Specify a Criterion 225
11.5 Step 3: Specify a Strategy 229
11.6 Step 4: Conduct the Analysis 233
11.7 Step 5: Evaluate the Reliability 233

11.8 Overall Implementation 238
11.9 Homework Exercises 297

12 Coding Schemes for Regression 299
12.1 Motivation 299
12.2 Reference Cell Coding 300
12.3 Cell Mean Coding 303
12.4 Classical ANOVA Coding 304
12.5 Effect Coding 306
12.6 Polynomial Coding 308
12.7 The Essence Matrix 309
12.8 Comments on Coding Schemes 310
12.9 Relationships among Coding Schemes 311
12.10 Homework Exercises 311

13 One-Way ANOVA 313
13.1 Motivation 313
13.2 Specification of the Model 313
13.3 (Usual) Overall Test 316
13.4 Defining and Estimating Cell Means 325
13.5 Which Means Are Different? 327
13.6 Contrasts 328
13.7 Conducting Multiple Comparisons 330
13.8 Homework Exercises 337

14 Complete, Two-Way Factorial ANOVA 339
14.1 Motivation 339
14.2 Model Concepts 340
14.3 Coding Schemes 342
14.4 Generating Cell Means 345
14.5 Computing Estimates and Tests 348
14.6 Contrast Matrices for Marginal Means 350
14.7 Choosing and Interpreting Tests 370
14.8 Step-Down Tests 373
14.9 Missing Data 380
14.10 Homework Exercises 382

15 Special Cases of Two-Way ANOVA and Random Effects Basics 385
15.1 Motivation 385
15.2 Blocking Variables and Block Designs 385
15.3 Fixed Block Design 386
15.4 Introduction to Random Effects Models 388
15.5 The Classical Approach to a Random Block Design 388
15.6 Role of Nonindependence of Observations 389
15.7 Computations for Mixed Models 390

15.8 Review Comments 390
15.9 Homework Exercises 399

16 The Full Model in Every Cell (ANCOVA as a Special Case) 401
16.1 Motivation 401
16.2 Cell Mean Style Coding of Full Model 402
16.3 Properties of the Model 403
16.4 Testing Strategies 409
16.5 Implementing Strategy 1, Adjusted ANOVA 410
16.6 Implementing Strategy 2, GLM Testing 411
16.7 Models of Interest 413
16.8 Implementing Strategy 3: Backwards Groupwise 415
16.9 Difference Scores: A Special Case of ANCOVA 417
16.10 Other Contrasts of Interest in ANCOVA 419
16.11 Other Tests of Interest in the Full Model 420
16.12 Regression (Reference Cell) Style Coding 421
16.13 Effect Style Coding 432
16.14 Modeling a Baseline Covariate 442
16.15 Comparing Coding Schemes 443
16.16 Generalizations 444
16.17 Homework Exercises 444

17 Understanding and Computing Power for the GLM 447
17.1 Motivation 447
17.2 GLM Theory 448
17.3 Factors in Choosing a Design 451
17.4 Using Parameter Estimates in Power Analysis 452
17.5 Example Power Analysis 452
17.6 Reporting Power 453
17.7 Tables and Software 456
17.8 How Much To Do? 457
17.9 Benefits 457
17.10 Homework Exercises 457

A Matrix Algebra for Linear Models 459
A.1 Basics 459
A.2 Matrix Properties and Decompositions 463
A.3 Principal Components (Basics) 469
A.4 Homework Exercises in Matrix Arithmetic 474
A.5 Homework Exercises for a Matrix Language 474

B Statistical Tables 477

C Study Guide for Linear Model Theory 489

D Homework and Example Data 495

E Introduction to SAS/IML 511

F A Brief Manual for LINMOD 513

G SAS/IML Power Program User's Guide 519

H Regression Model Selection Data 535

References 545

Index 551