

Contents

Acknowledgments xi

Chapter 1 Introduction

1.1 *About This Book* 1

1.2 *Statistical Topics and SAS Procedures* 1

Chapter 2 Regression

2.1 *Introduction* 3

2.2 *The REG Procedure* 4

2.2.1 *Using the REG Procedure to Fit a Model with One Independent Variable* 5

2.2.2 *The P, CLM, and CLI Options: Predicted Values and Confidence Limits* 10

2.2.3 *A Model with Several Independent Variables* 12

2.2.4 *The SSI and SS2 Options: Two Types of Sums of Squares* 14

2.2.5 *Tests of Subsets and Linear Combinations of Coefficients* 17

2.2.6 *Fitting Restricted Models: The RESTRICT Statement and NOINT Option* 18

2.2.7 *Exact Linear Dependency* 21

2.3 *The GLM Procedure* 22

2.3.1 *Using the GLM Procedure to Fit a Linear Regression Model* 22

2.3.2 *Using the CONTRAST Statement to Test Hypotheses about Regression Parameters* 24

2.3.3 *Using the ESTIMATE Statement to Estimate Linear Combinations of Parameters* 26

2.4 *Statistical Background* 27

2.4.1 *Terminology and Notation* 27

2.4.2 *Partitioning the Sums of Squares* 28

2.4.3 *Hypothesis Tests and Confidence Intervals* 29

2.4.4 *Using the Generalized Inverse* 31

Chapter 3 Analysis of Variance for Balanced Data

3.1 *Introduction* 33

3.2 *One- and Two-Sample Tests and Statistics* 34

3.2.1 *One-Sample Statistics* 34

3.2.2 *Two Related Samples* 37

3.2.3 *Two Independent Samples* 39

3.3 The Comparison of Several Means: Analysis of Variance 42

- 3.3.1 Terminology and Notation 42
 - 3.3.1.1 Crossed Classification and Interaction Sum of Squares 44
 - 3.3.1.2 Nested Effects and Nested Sum of Squares 45
- 3.3.2 Using the ANOVA and GLM Procedures 46
- 3.3.3 Multiple Comparisons and Preplanned Comparisons 48

3.4 The Analysis of One-Way Classification of Data 49

- 3.4.1 Computing the ANOVA Table 52
- 3.4.2 Computing Means, Multiple Comparisons of Means, and Confidence Intervals 55
- 3.4.3 Planned Comparisons for One-Way Classification: The CONTRAST Statement 56
- 3.4.4 Linear Combinations of Model Parameters 59
- 3.4.5 Testing Several Contrasts Simultaneously 59
- 3.4.6 Orthogonal Contrasts 60
- 3.4.7 Estimating Linear Combinations of Parameters: The ESTIMATE Statement 60

3.5 Randomized-Blocks Designs 62

- 3.5.1 Analysis of Variance for Randomized-Blocks Design 64
- 3.5.2 Additional Multiple Comparison Methods 65
- 3.5.3 Dunnett's Test to Compare Each Treatment to a Control 70

3.6 A Latin Square Design with Two Response Variables 72

3.7 A Two-Way Factorial Experiment 74

- 3.7.1 ANOVA for a Two-Way Factorial Experiment 75
- 3.7.2 Multiple Comparisons for a Factorial Experiment 78
- 3.7.3 Multiple Comparisons of METHOD Means by VARIETY 80
- 3.7.4 Planned Comparisons in a Two-Way Factorial Experiment 82
- 3.7.5 Simple Effect Comparisons 84
- 3.7.6 Main Effect Comparisons 85
- 3.7.7 Simultaneous Contrasts in Two-Way Classifications 86
- 3.7.8 Comparing Levels of One Factor within Subgroups of Levels of Another Factor 87
- 3.7.9 An Easier Way to Set Up CONTRAST and ESTIMATE Statements 89

Chapter 4 Analyzing Data with Random Effects

4.1 Introduction 91

4.2 Nested Classifications 93

- 4.2.1 Analysis of Variance for Nested Classifications 96
- 4.2.2 Computing Variances of Means from Nested Classifications and Deriving Optimum Sampling Plans 99
- 4.2.3 Analysis of Variance for Nested Classifications: Using Expected Mean Squares to Obtain Valid Tests of Hypotheses 99
- 4.2.4 Variance Component Estimation for Nested Classifications: Analysis Using PROC MIXED 101
- 4.2.5 Additional Analysis of Nested Classifications Using PROC MIXED: Overall Mean and Best Linear Unbiased Prediction 104

- 4.3 Blocked Designs with Random Blocks 106**
 - 4.3.1 *Random-Blocks Analysis Using PROC MIXED* 107
 - 4.3.2 *Differences between GLM and MIXED Randomized-Complete-Blocks Analysis: Fixed versus Random Blocks* 110
 - 4.3.2.1 *Treatment Means* 111
 - 4.3.2.2 *Treatment Differences* 112
- 4.4 The Two-Way Mixed Model 113**
 - 4.4.1 *Analysis of Variance for the Two-Way Mixed Model: Working with Expected Mean Squares to Obtain Valid Tests* 114
 - 4.4.2 *Standard Errors for the Two-Way Mixed Model: GLM versus MIXED* 117
 - 4.4.3 *More on Expected Mean Squares: Determining Quadratic Forms and Null Hypotheses for Fixed Effects* 120
- 4.5 A Classification with Both Crossed and Nested Effects 122**
 - 4.5.1 *Analysis of Variance for Crossed-Nested Classification* 124
 - 4.5.2 *Using Expected Mean Squares to Set Up Several Tests of Hypotheses for Crossed-Nested Classification* 124
 - 4.5.3 *Satterthwaite's Formula for Approximate Degrees of Freedom* 129
 - 4.5.4 *PROC MIXED Analysis of Crossed-Nested Classification* 131
- 4.6 Split-Plot Experiments 135**
 - 4.6.1 *A Standard Split-Plot Experiment* 136
 - 4.6.1.1 *Analysis of Variance Using PROC GLM* 137
 - 4.6.1.2 *Analysis with PROC MIXED* 139

Chapter 5 Unbalanced Data Analysis: Basic Methods

- 5.1 Introduction 141**
- 5.2 Applied Concepts of Analyzing Unbalanced Data 142**
 - 5.2.1 *ANOVA for Unbalanced Data* 144
 - 5.2.2 *Using the CONTRAST and ESTIMATE Statements with Unbalanced Data* 146
 - 5.2.3 *The LSMEANS Statement* 147
 - 5.2.4 *More on Comparing Means: Other Hypotheses and Types of Sums of Squares* 147
- 5.3 Issues Associated with Empty Cells 148**
 - 5.3.1 *The Effect of Empty Cells on Types of Sums of Squares* 149
 - 5.3.2 *The Effect of Empty Cells on CONTRAST, ESTIMATE, and LSMEANS Results* 150
- 5.4 Some Problems with Unbalanced Mixed-Model Data 151**
- 5.5 Using the GLM Procedure to Analyze Unbalanced Mixed-Model Data 152**
 - 5.5.1 *Approximate F-Statistics from ANOVA Mean Squares with Unbalanced Mixed-Model Data* 152
 - 5.5.2 *Using the CONTRAST, ESTIMATE, and LSMEANS Statements in GLM with Unbalanced Mixed-Model Data* 155

- 5.6 *Using the MIXED Procedure to Analyze Unbalanced Mixed-Model Data* 156
- 5.7 *Using the GLM and MIXED Procedures to Analyze Mixed-Model Data with Empty Cells* 158
- 5.8 *Summary and Conclusions about Using the GLM and MIXED Procedures to Analyze Unbalanced Mixed-Model Data* 161

Chapter 6 Understanding Linear Models Concepts

- 6.1 *Introduction* 163
- 6.2 *The Dummy-Variable Model* 164
 - 6.2.1 *The Simplest Case: A One-Way Classification* 164
 - 6.2.2 *Parameter Estimates for a One-Way Classification* 167
 - 6.2.3 *Using PROC GLM for Analysis of Variance* 170
 - 6.2.4 *Estimable Functions in a One-Way Classification* 175
- 6.3 *Two-Way Classification: Unbalanced Data* 179
 - 6.3.1 *General Considerations* 179
 - 6.3.2 *Sums of Squares Computed by PROC GLM* 182
 - 6.3.3 *Interpreting Sums of Squares in Reduction Notation* 183
 - 6.3.4 *Interpreting Sums of Squares in μ -Model Notation* 185
 - 6.3.5 *An Example of Unbalanced Two-Way Classification* 188
 - 6.3.6 *The MEANS, LSMEANS, CONTRAST, and ESTIMATE Statements in a Two-Way Layout* 191
 - 6.3.7 *Estimable Functions for a Two-Way Classification* 194
 - 6.3.7.1 *The General Form of Estimable Functions* 194
 - 6.3.7.2 *Interpreting Sums of Squares Using Estimable Functions* 196
 - 6.3.7.3 *Estimating Estimable Functions* 201
 - 6.3.7.4 *Interpreting LSMEANS, CONTRAST, and ESTIMATE Results Using Estimable Functions* 201
 - 6.3.8 *Empty Cells* 203
- 6.4 *Mixed-Model Issues* 214
 - 6.4.1 *Proper Error Terms* 214
 - 6.4.2 *More on Expected Mean Squares* 216
 - 6.4.3 *An Issue of Model Formulation Related to Expected Mean Squares* 221
- 6.5 *ANOVA Issues for Unbalanced Mixed Models* 222
 - 6.5.1 *Using Expected Mean Squares to Construct Approximate F-Tests for Fixed Effects* 222
- 6.6 *GLS and Likelihood Methodology Mixed Model* 225
 - 6.6.1 *An Overview of Generalized Least Squares Methodology* 225
 - 6.6.2 *Some Practical Issues about Generalized Least Squares Methodology* 227

Chapter 7 Analysis of Covariance

- 7.1 Introduction 229**
- 7.2 A One-Way Structure 230**
 - 7.2.1 Covariance Model 230
 - 7.2.2 Means and Least-Squares Means 234
 - 7.2.3 Contrasts 237
 - 7.2.4 Multiple Covariates 238
- 7.3 Unequal Slopes 239**
 - 7.3.1 Testing the Heterogeneity of Slopes 240
 - 7.3.2 Estimating Different Slopes 241
 - 7.3.3 Testing Treatment Differences with Unequal Slopes 244
- 7.4 A Two-Way Structure without Interaction 247**
- 7.5 A Two-Way Structure with Interaction 249**
- 7.6 Orthogonal Polynomials and Covariance Methods 256**
 - 7.6.1 A 2×3 Example 256
 - 7.6.2 Use of the IML ORPOL Function to Obtain Orthogonal Polynomial Contrast Coefficients 259
 - 7.6.3 Use of Analysis of Covariance to Compute ANOVA and Fit Regression 261

Chapter 8 Repeated-Measures Analysis

- 8.1 Introduction 265**
- 8.2 The Univariate ANOVA Method for Analyzing Repeated Measures 269**
 - 8.2.1 Using GLM to Perform Univariate ANOVA of Repeated-Measures Data 270
 - 8.2.2 The CONTRAST, ESTIMATE, and LSMEANS Statements in Univariate ANOVA of Repeated-Measures Data 272
- 8.3 Multivariate and Univariate Methods Based on Contrasts of the Repeated Measures 274**
 - 8.3.1 Univariate ANOVA of Repeated Measures at Each Time 274
 - 8.3.2 Using the REPEATED Statement in PROC GLM to Perform Multivariate Analysis of Repeated-Measures Data 275
 - 8.3.3 Univariate ANOVA of Contrasts of Repeated Measures 279
- 8.4 Mixed-Model Analysis of Repeated Measures 280**
 - 8.4.1 The Fixed-Effects Model and Related Considerations 281
 - 8.4.2 Selecting an Appropriate Covariance Model 284
 - 8.4.3 Reassessing the Covariance Structure with a Means Model Accounting for Baseline Measurement 291
 - 8.4.4 Information Criteria to Compare Covariance Models 292
 - 8.4.5 PROC MIXED Analysis of FEV1 Data 296
 - 8.4.6 Inference on the Treatment and Time Effects of FEV1 Data Using PROC MIXED 298
 - 8.4.6.1 Comparisons of DRUG*HOUR Means 299
 - 8.4.6.2 Comparisons Using Regression 301

Chapter 9 Multivariate Linear Models

- 9.1 Introduction 305**
- 9.2 A One-Way Multivariate Analysis of Variance 306**
- 9.3 Hotelling's T^2 Test 309**
- 9.4 A Two-Factor Factorial 312**
- 9.5 Multivariate Analysis of Covariance 317**
- 9.6 Contrasts in Multivariate Analyses 320**
- 9.7 Statistical Background 321**

Chapter 10 Generalized Linear Models

- 10.1 Introduction 325**
- 10.2 The Logistic and Probit Regression Models 328**
 - 10.2.1 Logistic Regression: The Challenger Shuttle O-Ring Data Example 328*
 - 10.2.2 Using the Inverse Link to Get the Predicted Probability 331*
 - 10.2.3 Alternative Logistic Regression Analysis Using 0-1 Data 334*
 - 10.2.4 An Alternative Link: Probit Regression 336*
- 10.3 Binomial Models for Analysis of Variance and Analysis of Covariance 339**
 - 10.3.1 Logistic ANOVA 339*
 - 10.3.2 The Analysis-of-Variance Model with a Probit Link 344*
 - 10.3.3 Logistic Analysis of Covariance 347*
- 10.4 Count Data and Overdispersion 353**
 - 10.4.1 An Insect Count Example 353*
 - 10.4.2 Model Checking 357*
 - 10.4.3 Correction for Overdispersion 362*
 - 10.4.4 Fitting a Negative Binomial Model 366*
 - 10.4.5 Using PROC GENMOD to Fit the Negative Binomial with a Log Link 367*
 - 10.4.6 Fitting the Negative Binomial with a Canonical Link 369*
 - 10.4.7 Advanced Application: A User-Supplied Program to Fit the Negative Binomial with a Canonical Link 372*
- 10.5 Generalized Linear Models with Repeated Measures—Generalized Estimating Equations 377**
 - 10.5.1 A Poisson Repeated-Measures Example 377*
 - 10.5.2 Using PROC GENMOD to Compute a GEE Analysis of Repeated Measures 379*

10.6	<i>Background Theory</i>	384
10.6.1	<i>The Generalized Linear Model Defined</i>	385
10.6.2	<i>How the GzLM's Parameters Are Estimated</i>	386
10.6.3	<i>Standard Errors and Test Statistics</i>	386
10.6.4	<i>Quasi-Likelihood</i>	387
10.6.5	<i>Repeated Measures and Generalized Estimating Equations</i>	388

Chapter 11 Examples of Special Applications

11.1	<i>Introduction</i>	389
11.2	<i>Confounding in a Factorial Experiment</i>	389
11.2.1	<i>Confounding with Blocks</i>	390
11.2.2	<i>A Fractional Factorial Example</i>	394
11.3	<i>A Balanced Incomplete-Blocks Design</i>	398
11.4	<i>A Crossover Design with Residual Effects</i>	402
11.5	<i>Models for Experiments with Qualitative and Quantitative Variables</i>	409
11.6	<i>A Lack-of-Fit Analysis</i>	413
11.7	<i>An Unbalanced Nested Structure</i>	416
11.8	<i>An Analysis of Multi-Location Data</i>	420
11.8.1	<i>An Analysis Assuming No Location×Treatment Interaction</i>	421
11.8.2	<i>A Fixed-Location Analysis with an Interaction</i>	423
11.8.3	<i>A Random-Location Analysis</i>	425
11.8.4	<i>Further Analysis of a Location×Treatment Interaction Using a Location Index</i>	428
11.9	<i>Absorbing Nesting Effects</i>	431

References	441
-------------------	------------

Index	447
--------------	------------

